



REFERENCE 1 (bases 1 to 1356)  
 AUTHORS Cao, J., Shapleigh, J., Gennis, R., Revzin, A. and Ferguson-Miller, S.  
 TITLE The gene encoding cytochrome c oxidase subunit II from Rhodospirillum rubrum: comparison of the deduced amino acid sequence with sequences of corresponding peptides from other species  
 JOURNAL Gene 101 (1), 133-137 (1991)  
 MEDLINE 91285423  
 PUBMED 1646008  
 REFERENCE 2 (bases 1 to 1356)  
 AUTHORS Hiser, C. and Ferguson-Miller, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUL-2001) Biochemistry Department, Michigan State University, East Lansing, MI 48824, USA  
 REMARK Sequence update by submitter  
 COMMENT On Jul 26, 2001 this sequence version replaced gi:15022158.  
 FEATURES  
 source  
 location/Qualifiers  
 1..1356  
 /organism="Rhodospirillum rubrum"  
 /db\_xref="taxon:1063"  
 349..1260  
 /gene="coxII"  
 349..1260  
 /gene="coxII"  
 /EC\_number="1.9.3.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome c oxidase subunit II"  
 /protein\_id="AA26100.3"  
 /db\_xref="GI:15022158"

BASE COUNT 239 a 447 c 437 g 233 t  
 ORIGIN

## alignment\_scores:

Quality: 165.00 Length: 42  
 Ratio: 4.125 Gaps: 0  
 Percent Similarity: 95.238 Percent Identity: 73.810

## alignment\_block:

US-09-712-768-4 x M57680 ..  
 Align seg 1/1 to: M57680 from: 1 to: 1356

```

1 Proleugluilevaltrphtlilevalprovalvalilleleuvalphe11 17
|||||
643 CCGCTCGAGTGCCTGCGAGATCGTCCGATCCTCCTCGGCAT 692
17 ecglyalapheserleuprovalleuphelysglnslnglnupheproglug 34
|||||
693 CGGGGCGCTTCTCGCTGCTGTTCAACGACGAGAAATCCCGGAG 742
34 lvsplilevalilleasnaValglugly 42
|||||
743 CGGACGTGACGAGTCAAGCTCAGCGGC 768
```

seq\_name: gb\_ba: PDCOX12

seq\_documentation\_block:

LOCUS PDCOX12 1653 bp DNA BCT 21-MAR-1995  
 DEFINITION P. denitrificans ox12 gene for cytochrome c oxidase subunit II (EC 1.9.3.1).  
 ACCESSION X05934  
 VERSION X05934.1 GI:45483  
 KEYWORDS cytochrome c oxidase subunit II.  
 SOURCE Paracoccus denitrificans.  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum rubrum group;

REFERENCE 1 (bases 1 to 1653)  
 AUTHORS Ludwig, B.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-AUG-1987) Ludwig B., Institut fuer Biochemie, Medizinische Universitaet zu Luebeck, Ratzeburger Allee 160, 2400 Luebeck 1  
 REFERENCE 2 (bases 1 to 1653)  
 AUTHORS Stehrucke, P., Steffens, G.C., Pankus, G., Buse, G. and Ludwig, B.  
 TITLE Subunit II of cytochrome c oxidase from Paracoccus denitrificans.  
 JOURNAL DNA sequence, gene expression and the protein  
 MEDLINE Eur. J. Biochem. 167 (3), 431-439 (1987)  
 PUBMED 88004464  
 FEATURES  
 source  
 location/Qualifiers  
 1..1653  
 /organism="Paracoccus denitrificans"  
 /strain="wild type"  
 /db\_xref="taxon:266"  
 /clone\_lib="PUC8"  
 /clone="470"  
 532..615  
 /product="signal peptide (AA -28 to -1)"  
 532..1425  
 /note="precursor polypeptide (AA -28 to 269)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA29372.1"  
 /db\_xref="GI:45484"  
 /db\_xref="SPTREMBL:059663"  
 /translation="MAITKRGVAAVNSLQVATMTVPALADVLGDLPLVIGKPVNG  
 GNFPDASSPLAHDOQMDHRYLTITAVTTFVCLLICTVRNRANVPARFTIN  
 TRLEIVTVPLILVAIGAFSLPLTFESQMPNDPLDIKAIQHOWMSYEPNDNF  
 ARDALWLEKEALADAGYSEDEYLLATIDMPVVPVGRKVLVQVATDVIHMTIIPAFV  
 KDDAVPGRITQALMFESVDOGVYFSGSELGINSNAVPIYKVAQSDEKYEAMLAGAKE  
 EFADASDYLLPASPVKLASAE"  
 616..1422  
 /product="mature cytochrome oxidase subunit II (AA 1-269)"

sig\_peptide  
 CDS

BASE COUNT 270 a 525 c 556 g 302 t  
 ORIGIN

## alignment\_scores:

Quality: 139.50 Length: 43  
 Ratio: 3.577 Gaps: 1  
 Percent Similarity: 90.698 Percent Identity: 60.465

## alignment\_block:

US-09-712-768-4 x PDCOX12 ..  
 Align seg 1/1 to: PDCOX12 from: 1 to: 1653

```

1 Proleugluilevaltrphtlilevalprovalvalilleleuvalphe11 17
|||||
841 CCGATGAGGTATGATCGACCTGCTCCGGTCTGATCTGCTGGCAT 890
17 ecglyalapheserleuprovalleuphelysglnslnglnupheproglug 33
|||||
891 CGGGGCGCTTCTCGCTGCTGTTCAACGACGAGAAATCCCGGAG 940
33 lvsplilevalilleasnaValglugly 42
|||||
941 ATCCGACCTGCTGATCAAGGCTCAGCGGC 969
```

seq\_name: gb\_ba: PDCOX1

seq\_documentation\_block:

LOCUS PDCOX1 4890 bp DNA BCT 12-SEP-1993  
 DEFINITION Paracoccus denitrificans cytochrome c oxidase CoII and CoIII genes  
 ACCESSION X05828  
 VERSION X05828.1 GI:45468  
 KEYWORDS CoII gene; cytochrome aa3; cytochrome c oxidase; unidentified reading frame.





```

seq_documentation_block:
LOCUS      AE006001      1188 bp      DNA      BCT      28-MAR-2001
DEFINITION Caulobacter crescentus section 327 of 359 of the complete genome.
ACCESSION  AE006001 AE005673
VERSION     AE006001.1 GI:13425117
KEYWORDS
SOURCE      Caulobacter crescentus.
ORGANISM    Caulobacter crescentus
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
Caulobacter.
1 (bases 1 to 1188)
Nieman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,
Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R.,
Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.,
Ely, B., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
Haff, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J.,
Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
Emolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and
Fraser, C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
TITLE       2 (bases 1 to 1188)
JOURNAL     Nieman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,
MEDLINE     Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I.,
AUTHORS     Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
Laub, M.T., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
Haff, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J.,
Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
Emolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
Fraser, C.M.
Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source      1. 11188
/organism="Caulobacter crescentus"
/db_xref="taxon:69394"
/complement(94..1752)
/gene="CC3406"
/complement(94..1752)
/gene="CC3406"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=1
/product="cytochrome c oxidase, subunit I"
/protein_id="AAK25368.1"
/db_xref="GI:13425118"
/translation="MASADTHDHAHDHDKPPFPAWFEFTNHKIDITLYIFALMA
GIYGGALSGILRMELMARGIOVSDTGMQKIFGKHGYNVTAHALIMFEVW
PAMIGFGNMFVPIIMIGAPDMAPRRNNISFWLLIAMILLVMEVDGPGHFGG
WTYIPLSITIGHKGPAMDALILSLHLAGASSIGAINETTFNNMRAPGMLHMPLE
VWSYLLAFLLLSIPVLAGAITMLLTRNFGTFPDPAGDPRVMOHLFMEFGE
VYILIPFGFIIISHIVTSFKRPVEGGLAMAVAMALIGVFVYVNAHMHWTGMSINL
RAYVAATMIIVPTGVKIFSMATMGSGISKRTKMLAIGTIFLFTGCVTGVLS
NAGIDYTHDTYVVAHFHVLISGAVFAIFAGFYIWEKEMCVKNEFLGAHFEWM
PIGVNLIFEPQHLGQGMRRYVDPAETLNNVSYVGMITTVVGVGMFLVLEA
AIRRKAENPWEAGATLEMTLSPPPHQFSEPPVYIKDDHH"
/complement(1779..2708)
/gene="CC3407"
/complement(1779..2708)
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=1
/product="cytochrome c oxidase, subunit II"
/protein_id="AAK25369.1"
/db_xref="GI:13425119"
/translation="KKAQFSGIRIRITGASALATMPEAGHALAEDLMCOPTPGADL
QPAASEKHAHMFHDMILMPLITATLLVLGLLWIVVYNNKKAAPVAKSSHNTI
EYIMTVVAVLILVAIFSRILFHNMDKPDLTAKATGYOMGYEYDOKRSII
SLPTKEQAEAKKVPYLLADNANMVPKRVVAVOVTAADVIHAFALPFGIKTAIP
GRINTEMKADRTGVYGCSELCGVADHAFMPEIENVYQAEFDANVYKAKTPAPAA
APAPAPASGATPASAPAPASVPAAPAPAA"
3083..3409
/gene="CC3408"
3083..3409
/gene="CC3408"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAK25370.1"
/db_xref="GI:13425120"
/translation="MPETIEIOLKGVIALCVIALSHADSYAYETASLTIDIMGE
GTIYPLMRMOSGLVETIVYESQSPKRYITRAGRESLAQKAEMTATRAVNN
IVAA"
3406..3969
/gene="CC3409"
3406..3969
/gene="CC3409"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK25371.1"
/db_xref="GI:13425121"
/translation="MTRAFETIRLRGSGIPATYADIADHEAFNSQAAGTGA
VLAALGDPDRLARERARAGLKKOMEKTPSNAAVITALIGLAIDVIFLPLMG
VLTGVLGLVAVYITTFVGGFIFAGAPLAAPGPTTAFITGILMASTFISLTL
VTVGNVAVVAVARLHFRLLKPLAEOA"
3966..4802
/gene="CC3410"
3966..4802
/gene="CC3410"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK25372.1"
/db_xref="GI:13425122"
/translation="MSRECTMLIRNTLVAGASALAVGCRAGVAAIAGELMTRK
TTPFGDSAVYTTDNGKIVTVRHVAGAGAPSPVWRTLPMTGKDTLLIDPVVVF
VGAERKIVSGPKSTVDRKAVGGRISLDDINLADGAILMTLGPGLQVRSRDRMKI
TVAPDVTRFEVKGSGDDIRRYEQALANLDSGVDVEAGKVTYVENSQGYAD
LGDLRKRAEALIEVSGSGAVPATGAKIDISGSDVELRTPKSVTSETIGSGVRR
EY"
4964..6436
/gene="CC3411"
4964..6436
/gene="CC3411"
/note="identified by match to PFAM protein family HMM
PF01523"
/codon_start=1
/transl_table=1
/product="ltd protein"
/protein_id="AAK25373.1"
/db_xref="GI:13425123"
/translation="MTSAFGLTCTCTNNAPTPASSLTLDAAGVDDPBARALICEALA
GADDELFERESSEAPFVDGRLKSAVDSGEGRLRVAGCTAGYAHAEISPAI
BRASASALAKRGVSAEGRATNEKLYEDNVSAPDSKVALQIDARAR
DPRVVOVNASMAGERVVEILRADGRLLRDVPLRVVQVYVERKDRREGSSGAG
RAGFAWLSIPKQDQVDEALRLMALVNDAAVACGAGDVVLGPMNGVLLHBAVHG
LEGDFNRKGISAFSGRIGERYAAPGVYFDDGSLJGRGSLTVDDGCTPTERTLLID
GLVGYMHDGMSARLMGKATNGRROSARHMPMRMTNGMLAGEHDPOMIASLR
GLYCANFGGGOVDITNGKFEVOCPEAYLVEGKITAPVKGATLIGDGPALTRVTMG
NDPDPGRTIGVCGKSGGVPGVQPSLKITGLTVGGTAV"
complement(6601..7098)
/gene="CC3412"
complement(6601..7098)
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=1
/product="acetyltransferase, GNAI family"
/protein_id="AAK25374.1"

```



```

CDS
    /gene="coxC"
    3314..4303
    /gene="coxC"
    /codon_start=1
    /transl_table=11
    /product="heme O synthase"
    /protein_id="CAA61745.1"
    /db_xref="GI:1177575"
    /db_xref="SPTREMBL:O51322"
    /translation="MCAHSCAHPTRKGSIVSVVDKAVELPRISAGVADYFALL
    KPRVSLVITFALVGLMTPGVHPVDFITALLAVGASGALMNALEDIDVLS
    KTRAPFIRGRTREGAFGLTFFSVMTGLVMVYAGLLAT11EFVYVLTG
    LKRTAQMIVIGGAGALPVVAMAATGSLVBLFLTFPFMPHPHVALAFPS
    DDYARAGVPLPVAGDPATRLQILLYIVLVAIAAAMPPLGFDMDIYGVTLTGAG
    MLVIAEYVRRHTGSOALRATRRRLFAFSIFLFLFALLVLDVAVAKAVPLIW"

BASE COUNT      749 a      1252 c      1415 g      964 t

alignment_scores:
    Quality: 99.00      Length: 42
    Ratio: 2.676      Gaps: 0
    Percent Similarity: 88.095      Percent Identity: 40.476

alignment_block:
    US-09-712-768-4 x NMCOXABC ..

Align seg 1/1 to: NMCOXABC from: 1 to: 4380

2 LeuGluLeuValrPrpHrileValProvalValIleLeuValPheIlecl 18
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1039 ATCGAGGTTGCGTGAGCATCATCCGTTGATCTTCCTGGCGCATTCG 1088
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
18 yAlaPheSerLeuProValLeuPheLysGlnGlnGlnPheProGluGlyA 35
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1089 GGYTCCCTCATTCCTGCTTTGTTTGAGCAGCTTGACCTTCGAGAGCGTG 1138
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
35 spIleValIleAsnValGluGlyA 43
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1139 ACCTGACGGTGAGGCGACCGGCAAG 1164
   ::::::::::::::::::::::::::::::::::::::::::::::::::::

seq_name: gb_ba:RPXX02

seq_documentation_block:
LOCUS      RPXX02      312430 bp      DNA      BCT      12-NOV-1998
DEFINITION Rickettsia prowazekii strain Madrid E, complete genome; segment
2/4.
ACCESSION  AJ235271 AJ235269
VERSION    AJ235271.1 GI:3868717
KEYWORDS   complete genome.
SOURCE     Rickettsia prowazekii.
ORGANISM   Rickettsia prowazekii.
REFERENCE  1 (bases 1 to 312430)
            Andersson, S.G., Zomrodipour, A., Andersson, J.O.,
            Eriksson, A.S., Winkler, H.H. and Kurland, C.G.
            The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
            Nature 396 (6707), 133-140 (1998)
TITLE      Nature 396 (6707), 133-140 (1998)
JOURNAL    99039499
MEDLINE    2 (bases 1 to 312430)
REFERENCE  Andersson, S.G.E.
            Direct Submission
            Submitted (11-NOV-1998) S.G.E. Andersson,
            Siv. Andersson Molbio. u.se, Dept. of Molecular Biology, University
            of Uppsala, Husaratan 3, Uppsala, S-751 24, SWEDEN
            On Nov 13, 1998 this sequence version replaced gi:3860788.
COMMENT    Location/Qualifiers
            1..312430
            /organism="Rickettsia prowazekii"
            /strain="Madrid E"

```

```

gene
    /db_xref="taxon:782"
    76..1830
    /gene="Rp226"
    76..1830
    /gene="Rp226"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="CAA1469.1"
    /translation="MLNKLCDLILFLINLLVTVQVYASPPPLPSPILAVPTDKN
    STNSISPEFKFQFQSKOKKMSOHOEOTKATHOESQIDSELENQSEPTI
    DPGSTLTPSAVNSIDSKAEYENSTLANSYQDIOVQOEEDPSEASEPIDIGTK
    FTSATNEMKEKAVSSNDKENTLISNITIPNPSPVISTPQADNVVVPVSOQVQYI
    KPLNLISRPDIPLNHTDILNKEKNELESTISMTTIPNNVSVPSIDPTOTLNT
    VSESESTKKQDWTYPIPVLYVDPNKSQSLALBOKNNNDIOINNOAEMHCVSSNV
    TIOKNDKVNNAATSESTKEFKVETQMFLPDDIVLCKLTEDATLEQMDHGYIKLF
    OKKEEMIANAERKRLVESFTKYDDINKKDIYANLSYSANDNAFRANERNNLELR
    ALLDVPPIQAKNSTGETLITSSITNGNTYLAKFLVIRKIKTSVLNDECKYPLIAL
    AKNTNIVCLMKAKGYN"
    2025..4013
    /gene="Rp227"
    2025..4013
    /gene="Rp227"
    /codon_start=1
    /transl_table=11
    /product="DNA GYRASE SUBUNIT B (gyrB)"
    /protein_id="CAA14690.1"
    /db_xref="GI:3860790"
    /translation="MSDLFSLNKKKKKIVTNTYSAKDIEVLGDEPYRKPRGMYIG
    TDSNAHHLVSEVLDNDEAVAGFASITIMKMDHSTITIDNKGRIPIDNRFPD
    KALAEVILTTLSGSKFSNNYVHTSGGLHGVISYVNAUSKFKIKVQKGLVSQSY
    SKGAKITDLSKASRLKGISINFTPDEIFSEKLNPKKTYEIAKSVLYKGV
    LMECEVEVSDIPKALINFPNGKLYLSKISLDNLVPIPEISGNIESTVDIKLE
    MAICWONDTSAPFMQSYCNTVPTPOGTHGGLSAIRGLAKYSEMGNKKAVALT
    EDILETASIVLSIFIVEPFOGOTEKEKLVSNVSKLVNIIKDHDSGLSDKVALTH
    LEHVAIAEFRNRKNERNISRSVTKIPGLACTPTASGCTFLFYEDSAG
    GSAKOARETROAVIPIMGKVLNVNASSFTLEKTIINNOAQDIEALACGSLNRYEN
    RYKTIIMTDADVDCAHIASLMTFFELMRPLVEGHLYLAKPLVLRQTSNNIYVA
    CDEEKIKLITVLSKASAKILEVGRFGIGEMMPALKEETMHPKRSILKVTLEDVO
    NVDKIVDILMGKPKPKRFQIYEQALVMDQIINKLID"
    4122..5447
    /gene="Rp228"
    4122..5447
    /gene="Rp228"
    /codon_start=1
    /transl_table=11
    /product="TAIL-SPECIFIC PROTEASE PRECURSOR (ctp)"
    /protein_id="CAA14691.1"
    /db_xref="GI:3860791"
    /translation="MLRFLTALPSTNCATIEGKETENKISNOQAYKQFOVFERIEK
    DYVQYPMQKIMDEAINGMNSLDPSHSYTYDDEDEDLEFTTKGEGGIVEIMYDSG
    AKIISISIDDLPAFRAGLKGDOYIVGVNDELVSTLGPKRAIKERKGPPTGKYLIIK
    BEEAPPOELELTRRELIVKIKPIKAHLEKNINVIIRITFNESTISLELAAYKLTESK
    DNLEKIIIDLRNMGILDOAIVASDYIDSGVITVTRGRTSSNSETKANEFSLAKP
    KVMPIVLINGNSASEIYVAGLLODHKATIIIGKRSKGSGVLOTOINSAAVKLT
    SKYVPPSRISOAGEIPEPTDIIIPAKVYVPEVKKIDKREPSSSLKNILKNMANKKS
    NKEETKNNKKESELSLKYKDDQFAFAYVITGLIINTVLETOGRK"
    5653..7146
    /gene="Rp229"
    5653..7146
    /gene="Rp229"
    /codon_start=1
    /transl_table=11
    /product="HISTIDINE KINASE SENSOR PROTEIN (bara)"
    /protein_id="CAA14692.1"
    /db_xref="GI:3860792"
    /translation="MNNKTYIVRFLSITFLINIVNMFLRYFMKEIMIKOYALEHT
    KIVELYDNINWIKHNVSKLHKFDYKLLODOOPITVITVITQMPFNINISYVL
    KGNKRTSNMKNHYSVNYKDDSLIEYITKIDKIFLKSPTSAAPLADAKGITSHTL
    LPKVIENESDLTIEHASFVSYTIPVDIHDNDFVDVAVFETINNTISQMNNTISLECK
    VFITFIITFIETIILISNTNYARQIIEQOLETRNLKAOIEKVEKTSSTNTFFANI

```



## FEATURES

gene

CDS

CDS

gene

sig\_peptide  
CDS

CDS

ene  
DS

DS

gene  
CDS

CDS

gene

CDS

CDS

CDS

ene

DS

5

ne

```

VFEPDHOAGLSGMRRRVYDYPDAFACMNLVSSVGSYISQFVLIPLXCVIDAFKAPV
AGDNDPAGAGATLETWLTLPSPPHQFVEVLRVO"
4137. 5081
/gene="coxh"
4137. 5081
/gene="coxh"
/function="protoheme ix farnesyltransferase"
/codon_start=1
/transl_table=11
/product="putative heme O synthase"
/protein_id="CAB56820.1"
/db_xref="GI:6006414"
/translacion="MSVLDQNNVDINPRISAEVGDYIALIKPVMSTVETFAALVMA
MAHGHNPVALATISLCTIAGVAGAGALNMLDEDDIADKSRINRNPITGRITRELA
LPPVGMALAVSTWMTGIIIVNNTAGALAFETFEYVYIYMKLRWQIINLTGGAGA
LRDRIILFSTYTLVILAVAAAPALGATPVATGVSTILTGAGMVLAIINVMRERSOS
5078. 5251
/gene="coxI"
5078. 5251
/gene="coxI"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56882.1"
/db_xref="GI:6006415"
/translacion="MIVADKPEPDGIVITLQADKRSRORSIALIALGVLVLEFV
TMVKRPALVYRM"
5257. 5901
/gene="coxG"
5257. 5901
/gene="coxG"
/function="biogenesis of cytochrome c oxidase"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="putative heme A synthase or assembly protein"
/protein_id="CAB56842.1"
/db_xref="GI:6006416"
/translacion="MNEPTIIRSDOSRIARKQVSGKIGLRDVLASTGCGVALMGAS
YAAPFQWPCFATGSGTQVATAPATGAPRIAKLIANDRSNAPQLPKFEPDSE
REMPVVEYDVSIAIDHENDGISTLTSTFTFPVKDPPVPLASGEDDKRKGMIL"
5981. 6877
/gene="coxC"
5981. 6877
/gene="coxC"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/product="cytochrome c oxidase subunit III"
/protein_id="CAB56882.1"
/db_xref="GI:6006417"
/translacion="MAQAQGNHDNYLVDPSRPWAVGSLSAFLMAVATAMNNHNSA
ATVGVGVSTVLEVTNMSWMDVLEKQYKGDHETRYVOLHNHNGMILPASTVMEFA
MWAFAFNALREPADAVNATRDVAVGGLGTADAGACAVAGPTWRPHGIEFTDPNNLEFLN
TLILTLSTQVTAHNAHLLENDROGKYGLILVYUGALFTCVQATVSHAAFSFVAG
LVGAFFENATGFGHFNVLGVGFIPLVCLPRATVAGHTPRQNHGFEFAVMYHFDVVW
LFLFCITVYVHGAGETMAHGNH"
6975. 7376
/ncote="ORF13"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56824.1"
/db_xref="GI:6006418"
/translacion="MNDTPASPKSETTVYIQAAMRCLACKSPRCGQGLVAGFLTLAS
CDRCGLATIDMGDSRAPIITMLAGVIVACALIVVEKYPYVMLHAVIMLPLILAT
LILPRAMKSLTILADPETHHKAAGRLVDRAK"
7199. 8121
/gene="shd1"

```





/protein\_id="AAAF78798.1"  
 /db\_xref="GI:8708907"  
 /translation="MPCGRIADVPPAPRAGGEGDESDVYINIMAFSGAGGCGCQQR  
 SGCRIARARRRRPRLQHDADPRGAGSTPAARVDLHQRRLGLLGARRACRAR  
 RRRPAPDRSREAEGTARGDITGVKIDIALPKRSLPLKADKSPRAITDDRPAAERPG  
 RSCSDVPEKGLASADRRHQPRPAREGVEDLSLSWRKPGHIRGLSVHSRALPPAA  
 VTRIGKNSWLPCATAIH"  
 8893..9240  
 /gene="uncI"  
 /note="orf115"  
 8893..11493  
 /gene="Fo ATPase operon"  
 /note="constitutively expressed"  
 8893..9240  
 /gene="uncI"  
 /note="ORF115; initial gene of the Fo ATPase operon"  
 /codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="unknown"  
 /protein\_id="AAAF78799.1"  
 /db\_xref="GI:8708908"  
 /translation="MEIADIRPVRVPEPNSSEILISCPNPEAATSPPKSPAGDDGR  
 AARASAMALGRISSEIVAGVVGAGIGWCPDRLLSPPGFLVFLILGVAGVNVV

alignment\_scores:  
 Quality: 92.00 Length: 42  
 Ratio: 2.629 Gaps: 0  
 Percent Similarity: 83.33 Percent Identity: 40.476

alignment\_block:  
 US-09-712-768-4 x BJU33883/rev ..

Align seg 1/1 to reverse of: BJU33883 from: 1 to: 31495

2 LeuGluIleValItrPThrIleValProValValIleLeuValPheIleG1 18  
 :::  
 25233 ATCGAGGTGGCCGTGACGCTGGTCCGGCTGATCGTCGTCGCATCTC 25184

18 yAlaPheSerIeuProValIleuPheLysIleuIleuPheProGluGlyA 35  
 :::  
 25183 GGTGGCGGTGTCGCCGCTCTCTCTCGAGCTCGACGTCGCCGAAGGGCGG 25134

35 sPlIleValIleAsnValGluGlyArg 43  
 :::  
 25133 ACCTGACCATCAAGGCACGCCGCAAG 25108

seq\_name: gb\_sy:SYNBCTCONS  
 seq\_documentation\_block:  
 LOCUS SYNBCTCONS 1260 bp DNA SYN 12-JUN-1993  
 DEFINITION Bacterial cytochrome oxidase subunit II consensus sequence, 5' end.  
 ACCESSION L07491  
 KEYWORDS L07491.1 GI:295264 cytochrome oxidase II.  
 SOURCE Artificial gene DNA.  
 ORGANISM synthetic construct  
 artificial sequence.  
 1 (bases 1 to 1260)  
 Arredondo-Peter,R., Escamilla,E., Helman-Meneguzzi,F. and Hernandez,F.  
 Computer analysis and modelling of a consensus sequence of bacterial cytochrome oxidase subunit II unpublished (1992)  
 location/Qualifiers  
 1..1260 location="synthetic construct"  
 /db\_xref="taxon:32630"  
 1..1260 /gene="coxIcs"  
 1..>1260 /gene="coxIcs"  
 /gene="coxIcs"  
 /GC\_content=41.0



```

/transl_table=4
/product="cytochrome oxidase subunit III"
/protein_id="AAC99649.1"
/db_xref="GI:4091915"
/translation="MSKTYHRYHLVDSPMPYIGSGKALLITGVSLEYFYMYIWMY
HGVITITITMYWMDIITREAFEGHHQVKGRLRGMLITSECEFFAFMAEF
HSSLSPTIELGAWPVPVIGIEVLDPFVPLNTALITSSGATVMAHIAIYSGRLAI
OSLCTVILGVQFALQAMEYEAPFISDSYGSSTFPMATGFHGVIIIGTIFELVC
LARLGYHYTHHHHGGFEAASWYHFDVAVMLFLVYCIWMSG"
/complement(6444..7157)
/gene="ATPase 6"
/complement(6444..7157)
/gene="ATPase 6"
/translation="MMAASFDFQFNVIRIGITITNSSIMTIFESITLIGCSYKL
IPRPLAOIEILYHNFELGLVKDSTANKQYFALLITITFTAGLMLGLFVYFPT
AHITITGLSLISLITAVTILGITQYRNFNSKMMPSGADPLLAPLIVSEVYSRA
ISLGIRLAANVSAGHLFAILASGFRAMISKMLVSLAPILVMIEITILEIAVALIO
AYVCLITITVINDSLNH"
/complement(7170..7388)
/gene="ATPase 8"
/complement(7170..7388)
/gene="ATPase 8"
/codon_start=1
/transl_table=4
/product="ATP synthetase subunit 6"
/protein_id="AAC99652.1"
/db_xref="GI:4091918"
/translation="MMAASFDFQFNVIRIGITITNSSIMTIFESITLIGCSYKL
IPRPLAOIEILYHNFELGLVKDSTANKQYFALLITITFTAGLMLGLFVYFPT
AHITITGLSLISLITAVTILGITQYRNFNSKMMPSGADPLLAPLIVSEVYSRA
ISLGIRLAANVSAGHLFAILASGFRAMISKMLVSLAPILVMIEITILEIAVALIO
AYVCLITITVINDSLNH"
/complement(7170..7388)
/gene="ATPase 8"
/complement(7170..7388)
/gene="ATPase 8"
/codon_start=1
/transl_table=4
/product="ATP synthetase subunit 8"
/protein_id="AAC99655.1"
/db_xref="GI:4091921"
/translation="MMPHLDITAVLYLOYSWTITLILALYSIMSLFLEPKIONLRIRS
ILOEPPAPYKQOPANSAYNIYDILKR"
/complement(7406..8167)
/gene="COII"
/complement(7406..8167)
/gene="COII"
/codon_start=1
/transl_table=4
/product="cytochrome oxidase subunit II"
/protein_id="AAC99650.1"
/db_xref="GI:4091916"
/translation="MKPTINALLEASLRLDAEPFQSFODASPVEEITILLINOIM
FLIITITVIMLIRGLTGOAYHRYLEGVITIEVMTIIPALIVFIAPPSKLITL
MDEVVEGTVKAIQHWYMSYSDYOTLIGESTLEPDSYMTPTSDLPDGLRIIE
VDNRWVPIIDTVRYLVGVADYLSFAFPAISLATMDAVPGRLNQTGFLAKRPGLEFGQ
CSELGANSHEMPYIEAVSMKTYISMLSSICADL"
/8279..9872
/gene="COI"
/8279..9872
/gene="COI"
/translation="TAA stop codon is completed by addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except="(pos:9872,aa:TERM)
/transl_table=4
/product="cytochrome oxidase subunit I"
/protein_id="AAC99653.1"
/db_xref="GI:4091919"
/translation="MNYLITRRLFTSTNHRDITGLYILFGAFSGMAGTASSMLIRELS
APSGMIGDDHLYNYIYSHALLMIFLVMMVMMGFGNMFVPLIMIGADMAPPRLNI
LSNITFTTLYNNKRVPMQSMNRPLFLVWSVLTITILLISLPLACATIMLLDRNEN
TTFPDAGGDDPLIFGLFWEFGHPEVYIILPGFMYSGOIIPTFSAKOHIGYGMV
YAMISICVGLFIYVAHHMTVMGDVDRATVATMTITVAPGIRKTFESLATTYGSF
RLDTPMLAIGFELFTIGLGVILANSLOIALDITYVVAHRYVLSKAVGAIF
GGYVYMGKITGVSYNELYKGIHFMIMFGVNLTFEPQHLGLAGLPARYSDPAYQ
DMNIVSCGAVIALVGIIMFTIYLYSTALAAERFRWSTSPGLEMSLSPPAFTYD
ELPEVYORKLSRDLNITSTLL"
/9866..10886
/product="small subunit ribosomal RNA"
/10927..>11715

```

```

CDS
/gene="ND1"
/10927..>11715
/gene="ND1"
/codon_start=1
/transl_table=4
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC99654.1"
/db_xref="GI:4091920"
/translation="MTRMSTIISIFKTIATIIPLVLAIVTLTAERVLYGQWARK
GNVGVYGLIOPADCKIKFTKEMAIIPNANLSVTVAPILSLTFLAMGVIFESP
GIVLADINCVLTITFALISSMGVAILMSGWSNRYAFGAIRAAQAMISYECIGLI
LISYILCVGSUNITQIVLAQAEIWTIIPFPALMFPASALAEITNAPFDLREGSEL
VSGYVYESSMSFALFELAEYGHIIIMSLCISILFLGWAAPTGT"
BASE COUNT 3356 a 1954 c 2333 g 4072 t
ORIGIN

alignment_scores:
Quality: 90.00 Length: 41
Ratio: 2.812 Gaps: 0
Percent Similarity: 78.049 Percent Identity: 43.902

alignment_block:
US-09-712-768-4 x AF064823/rev ..
Align seg 1/1 to reverse of: AF064823 from: 1 to: 11715

2 LeuGIUlleValTTPThIleValProValValIleuValPheIleGI 18
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7939 AVAGAGATGTTGACATPAAATCCAGCAATATATTAGTGTTATAGC 7890
18 yAlaPheSerIeuProValIleuPheLysGInGInGluPheProGluGI 35
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7889 ATTCCTTCTCTGAATTAATTACTTATTTGANGAGAGAGTAGTAGACCG 7840
35 spIleValIleAsnValGIuGI 42
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7839 GTGTACAGTAAGCAACCATAGGT 7817

seq_name: gb_in:KTU09810
seq_documentation_block:
LOCUS KTU09810 15532 bp DNA circular INV 22-JUL-1998
DEFINITION Katharina tunicata mitochondrial genome, complete sequence.
ACCESSION U09810
VERSION U09810.1 GI:557273
KEYWORDS
SOURCE
ORGANISM
black chiton.
Mitochondrion Katharina tunicata
Eukaryota; Metazoa; Mollusca; Polyplacophora; Neoloricata;
Ischnochitonina; Mopaliidae; Katharina.
REFERENCE
1 (bases 1 to 15532)
Boore,J.L. and Brown,W.M.
Complete DNA sequence of the mitochondrial genome of the black
chiton, Katharina tunicata
Genetics 138 (2), 423-443 (1994)
JOURNAL
Genetics 138 (2), 423-443 (1994)
MEDLINE
95129806
REFERENCE
2 (bases 1 to 15532)
Boore,J.L.
Direct Submission
Submitted (17-MAY-1994) Jeffrey L. Boore, Cell Biology and
Neuroanatomy, University of Minnesota, 321 Church St. SE,
Minneapolis, MN 55455, USA
FEATURES
source
location/Qualifiers
1..15532
/organism="Katharina tunicata"
/organelle="mitochondrion"
/db_xref="taxon:34587"
1..1548
/note="followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/codon=(seq:"tga", aa:"TTP")
CDS

```

```

/product="cytochrome oxidase subunit I"
/protein_id="AAC48362.1"
/db_xref="GI:557274"
/translation="MRWIFSTNKHIDIGTLYILFGIWAGLVGALSLILBAELGQPCAL
LGDDQYANVIVTAHAFVMIFFLVPMWIGFGNMVLPMGLGVDMAPRINNMSEFLL
PPLACILLAGAVESGAGTGWIVPPPLAGVAGAGSDVLAIFSLIAGVSSILGAVN
FIVIVYMRSEGMOLERLPLFVMSVITALLILILPYLAGGITMLITBRNEMTFED
PAGGGPILYQHLFFRFGHREYVILIPGGMISHIYMHXSSKKEFGGLKMYALIA
IGLGLIVMAHMEVVGMDVTRAFYFAAMITAVPGIKIESWLTITGARIGSETP
MYMALGFIPLFTVGLGLIVLSSSLDLMHDSYVVAHFVYLSMGAVFALFAGNV
WYPLMGLSLHERMTKSHFVWFLGYNLTFEPHFLGLSGMPRRYSDDYPCVYIKWNY
SSIGSLSEFAVFLFPMFIVESIVSQRVIMSSLSALEWDSIDPLGFLDKKGLL
FVNLILL"
1619. 1682
/feature="codon recognized: GAC"
/product="tRNA-Asp"
/anticodon="(pos:1650..1652,aa:Asp)
2107. 2796
/codon_start=1
/translation=1
/codon=1
/product="(seq:"lga",aa:Trp)
/product="cytochrome oxidase subunit II"
/protein_id="AAC48365.1"
/db_xref="GI:557277"
/translation="MAFWSQMGQDQASPIEMEDLIFPHDAMILLMIISLSYGAVS
LMNSFSLRSFSLSEQIEIYVTLIPAVVLIPLAFLPSLQLLIDDELEPALTIKVGH
QWVSYSPSDFINERDSYMSLEDEEGGYRLLENHRSVPMKTRVRLVYAADVL
HSMVPSLGVKADVPGRNLQSLFFANPPGVFFGQCSFICGANSHPVILEVVDSS
FIRKIMFNGEA"
2872. 3033
/codon_start=1
/translation_except="(pos:3001..3003,aa:Trp)
/translation=5
/product="ATPase subunit 8"
/protein_id="AAC48364.1"
/db_xref="GI:3335323"
/translation="MPQLAPMNIIFLLIFWNSVFCVGVVWVWVSSSFLSKSVSW
PEYFKWVW"
3037. 3729
/codon_start=1
/translation=5
/product="ATPase subunit 6"
/protein_id="AAC48363.1"
/db_xref="GI:557275"
/translation="MMMDIFSSPDNNPTLFSFYLIMVGLAPVLISQNMWESKSR
FYSMLPFKNMGQVWRAATKNSGRTNWTISFLLIIVNLSGLIPYPSLSSHS
FACGFLPFWLSLIFSGLVSEVSHLPSGAGIILPFLVLEVETYSISRPITLESY
RLAANMSAGHIIIGLISYLSVGLFCYPIILVGLLFLVQVFFLFEIGVSLQVIFS
LLVLSYSDHRY"
complement(3758..3822)
/feature="codon recognized: UUC"
/product="tRNA-Phe"
/anticodon="(pos:3789..3791,aa:Phe)
complement(3821..5536)
/feature="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAC48366.1"
/db_xref="GI:557278"
/translation="WKMMFLSTQATVILLSHSLSESTAMTLLTNNKILILEFCFYN
FNSVITLPIILNTTALFSAIVCYISASVMIPTKNYSODIFIKRFVIMLVLS
NLLIFPLIATLLGMDGLVSCFVLYONYKSLGALMTAFMNRIGDAVILLAIS
VTFSGHNALEFMSTPTMTLLISMIILAGTKSQLPFCMLPAMAPAPVAVLAV
SSTIYVAGVPIILIOYPPILNSHLSTSLMISSWLMAGISLANFEDDKIIALST
ISQGVMAISVSLGLFSLALPHLETHAFKALPLFCAGNIIISHNNNDIKMSHWI
QMPESSTFNANLALGCPPEWAGYSKDVITEMMENONNELISLANFATITLSTAS
SARISLIFWSKNTQSSVSTSDENSLIAPLTMAGAITSGAVFSLILTPSTLP
ISMNKLATFATITGAATYKLFLEKLIKSHSIINPFPMWEMINLSNMNKSSSL
YTGLEYFSLDGMWETIGEGILKIMTLTINOKNOSNLFNFIISLILSMLEISV
TWMP"
complement(5537..5601)
/feature="codon recognized: CAC"

```

```

/product="tRNA-His"
/anticodon="(pos:5568..5570,aa:His)
complement(5597..6928)
/feature="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation_except="(pos:complement(6788..6790),aa:Trp)
/translation=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAC48367.1"
/db_xref="GI:3335324"
/translation="WSAFSTVTKSKNWKMSISHPFMTITLNFSPMNPKN
YMFEDNSVPLMLISCWITGLMSSYKILKNNLVSEFLNWMILNFIITWPTOK
SLFSLIFFEASLIEPLIILIMMGYQPPRLQAGMYMTITIGALPLILNIFLYSHN
AHLNLLIMSLPIEYQAMISFWMLFVILFVLPKPIYSFHLMLFKHAEVAGSMI
LAAILIKIGVGINIINNIPKYNVFLVFTITLGLVGVLSLCLQSDMKALIAV
SSVGHMGLMAGLISSEFGKLMALIMIAHGLSSGLFSSNMWYEGSGRSLETK
GVTLSVPTFSKSFILMCSINMAAPPSINLSAGIMISTVTSKRLMLLMAFVSA
VITLPLFNTQHKISSFVNPNNFRLYVNIPLMWIPADLILITTTTC"
complement(6940..7242)
/codon_start=1
/translation=5
/product="NADH dehydrogenase subunit 4L"
/protein_id="AAC48368.1"
/db_xref="GI:557280"
/translation="MKNFNPPLFTGLAUSISALMTICLQKRLNLTLLAEIIMVSI
FLLVSLISNMISNEGIMFVLLTLAASEASIGLAMLVILIRAHGMDVSSISITHC"
7247. 7312
/feature="codon recognized: ACA"
/product="tRNA-Thr"
/anticodon="(pos:7275..7277,aa:Thr)
complement(7312..7378)
/feature="codon recognized: UCA"
/product="tRNA-Ser"
/anticodon="(pos:7348..7350,aa:Ser)
complement(7381..8520)
/codon_start=1
/translation_except="(pos:complement(8176..8178),aa:Trp)
/translation=5
/product="cytochrome b"
/protein_id="AAC48371.1"
/db_xref="GI:3335325"
/translation="MLKPRSSHVPKIKISLIDPSPNSLIMNNGSLIGLIL
QILTLFLAMHTATNIELAFSSVAHICRQVNGMLIRAHANGSLFICLYFHLGRG
MYGGSFVNIHTWVGVILFLTWGTAFTCVLPFGQMSFGWATITNLSAIPYIGTT
MYOMINGGFVADNATLRFESFHLAPFIIAASVAILFHEGSSNNLGLINSSEK
VTHSHPTKLDVGFVWVYFLFLVFLDPLNLAHDNFIIPANPLVTPHIDPEWFL
FAVAILRSIPKRGVGVGLMSIVLFEVPLINLKGFTISFVPTQILYMSLVSVL
LTMVGAQVDEPFIILGOTLPFVFLFLMLPDKKMDWTIN"
complement(8513..9019)
/feature="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation=5
/product="NADH dehydrogenase subunit 6"
/protein_id="AAC48369.1"
/db_xref="GI:557281"
/translation="MPLFLMSLIFSCASPLILTOPISGLILFMSLPISTLIFPG
SSTWFSPIELIYIGLIVFAVYVTALEPMKRNKLIQVFPVSTFVWLLMILISDF
IGILLSEOTKMFLLMSFFNHLGISLSPENLIIIVSLALILEVFLICVVKICYENG
PLRPKYA"
9024. 9089
/feature="codon recognized: CCA"
/product="tRNA-Pro"
/anticodon="(pos:9054..9056,aa:Pro)
complement(9082..10032)
/codon_start=1
/translation=5
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC48370.1"
/db_xref="GI:557282"
/translation="MPLSFSWVYIAYISILLAAVFTLLERKGLGTYORRKGKXYSI
GGLAOPLADALKLFLEKNNIYNKSLPYTAVLSLISLCLMSLYWSDFTLWTFE
GVLEPCTALSIVGTGIMAGWASNKALRAVAVQIISYEVTLFLVLLSAAALS
SLNINQSSHQLLYWNSILWPIILMIIVYCAETNRAAPPDAEGSEIVSGNIEYG

```





```

/product="cytochrome oxidase subunit II"
/db_xref="GI:14278905"
/translation="MAHPTQUGFODASPIEELLHFHDHLMVALLISMLVLIISV
LMTTTLSSSTNIDAEIEMVMTIMPAIILIVALLPSRLILMDEISPDVITIGH
OMWMSYSEDFDLSEFDSYMIPTKDLPEHRLLEVDNMCWTPIGTARILTPMEDVL
HSMVAVPALGVKSDAIPGRLNOSFLIAHPGIYGGCSEICGANSFMPIVEALPVRN
FLNWIPTAODA"
10330..10398
/note="codon recognized: AAA"
/product="tRNA-Lys"
/anticodon="(pos:10360..10362,aa:Lys)"
10399..10563
/gene="Atp8"
10399..10563
/gene="Atp8"
/codon_start=1
/translation="IIEFHQIWEILITITAVLMLIIRSLTTRHYHRLIDCTLIET
IWTLIPAVLIVETAPPSLKLILYIMDEVIDPALTIKAYIGHQWWSYEYSDYGTIDFED
SYMIPPTDLSGDLRLLEVDNRYSSALFKLSIYV"
/db_xref="GI:14278906"
/translation="MPQLNPSFWLFLFLMLILITFLAPSKILIGHTNLNEPNKGTWT
TNTFTWTPWQ"
10534..11235
/gene="Atp6"
10534..11235
/gene="Atp6"
10554..11235

alignment_scores:
  Quality: 84.00      Length: 41
  Ratio: 2.625      Gaps: 0
  Percent Similarity: 78.049      Percent Identity: 41.463

alignment_block:
  US-09-712-768-4 x AB043889 ..
  Align seg 1/1 to: AB043889 from: 1 to: 17804

      2 LeuGIIIEValITPThrIIeValProValValIIleuValPheIIegI 18
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      9822 ATTGAATATGCTCGACAAATATGCACGCTATTAATCTAATGCTAATGCG 9871
      18 yAlaPheSerLeuProValIleuPheLysGInGluPheProGluGlyA 35
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      9872 CCTTCATCCCTCCGATTCCTACTTAATAGACGAATTAAGCAATCCAG 9921
      35 sPILeValIIeAsnValGluGly 42
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      9922 ACATAACCGTAATAACAATTGCT 9944

seq_name: gb_in:AF362018

seq_documentation_block:
  LOCUS AF362018 416 bp DNA INV 28-MAR-2001
  DEFINITION Clathrina sp. RFW-2001 cytochrome c oxidase subunit II (COII) gene,
  partial cds; mitochondrial gene for mitochondrial product.
  ACCESSION AF362018
  VERSION AF362018.1 GI:13469937
  KEYWORDS
  SOURCE
  ORGANISM
    Clathrina sp. RFW-2001.
    Mitochondrion Clathrina sp. RFW-2001
    Eukaryota; Metazoa; Porifera; Calcarea; Calcinea; Clathrinida;
    Clathrinidae; Clathrina.
  1 (bases 1 to 416)
  Watkins, R.F. and Beckenbach, A.T.
  The rate of sequence evolution in mitochondrial genes of the
  Demospongiae, Calcispongiae, and Anthozoa is very low relative to
  the bilateral animals
  Unpublished
  2 (bases 1 to 416)
  REFERENCES
    Watkins, R.F., Stewart, J.B. and Beckenbach, A.T.
    Direct Submission
    Submitted (19-MAR-2001) Department of Molecular Biology and
    Biochemistry, Simon Fraser University, 8888 University Drive,

```

```

FEATURES
  source
    Burnaby, BC V5A 1S6, Canada
    Location/Qualifiers
      1..416
      /organism="Clathrina sp. RFW-2001"
      /organellar="mitochondrion"
      /db_xref="taxon:135677"
      <1..>416
      /gene="COII"
      <1..>416
      /gene="COII"
      /codon_start=3
      /translation="IIEFHQIWEILITITAVLMLIIRSLTTRHYHRLIDCTLIET
      IWTLIPAVLIVETAPPSLKLILYIMDEVIDPALTIKAYIGHQWWSYEYSDYGTIDFED
      SYMIPPTDLSGDLRLLEVDNRYSSALFKLSIYV"
  CDS
    BASE COUNT 135 a 50 c 75 g 135 t 1 others
    ORIGIN

alignment_scores:
  Quality: 83.00      Length: 41
  Ratio: 2.594      Gaps: 0
  Percent Similarity: 78.049      Percent Identity: 41.463

alignment_block:
  US-09-712-768-4 x AF362018 ..
  Align seg 1/1 to: AF362018 from: 1 to: 416

      2 LeuGIIIEValITPThrIIeValProValValIIleuValPheIIegI 18
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      126 APTGAATATGCTCGACCTTAATCCGCACTTATTAATGATTAATGATTAATGAC 175
      18 yAlaPheSerLeuProValIleuPheLysGInGluPheProGluGlyA 35
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      176 GTTCCTTCGTTAAATATATGTTAATGATTAATGATTAATGATTAATGACCCGG 225
      35 sPILeValIIeAsnValGluGly 42
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      226 CCTTAAGTATTAAGCTAATAGGG 248

```



• •  
• Thu Dec 27 08:21:50 2001

us-09-712-768-4.rge

---









```

FT      misc_feature      /tag= m
FT      /note= "Oligonucleotide 3R"
FT      complement (3739..3723)
FT      /tag= n
FT      /note= "Oligonucleotide 7R"
FT      4354..4338
FT      /tag= o
FT      /note= "Oligonucleotide 2R"
FT      complement (4590..4574)
FT      /tag= p
FT      /note= "Oligonucleotide 2F"
FT      4705..4721
FT      /tag= q
FT      /note= "Oligonucleotide 1R"
FT      5030..5046
FT      /tag= t
FT      /note= "Oligonucleotide 9P"
FT      complement (5315..5299)
FT      /tag= s
FT      /note= "Oligonucleotide 1F"
FT      6017..6033
FT      /tag= l
FT      /note= "Oligonucleotide 6R"
FT      complement (6164..6148)
FT      /tag= u
FT      /note= "Oligonucleotide 6F"

XX      W09119005-A.
XX      PD      12-DEC-1991.
XX      PF      31-MAY-1991; 91WO-GB00869.
XX      PR      01-JUN-1990; 90GB-0012196.
XX      PA      (ISIS-) ISIS INNOVATION LTD.
XX      PI      Wakefield AE, Hopkin JM, Moxon ER;
XX      WPI: 1992-007487/01.
XX      DR      P-PSDB; AAR20056, AAR21409, AAR21410, AAR21411, AAR21412, AAR21413.
XX      PT      New DNA sequences which act as oligo:nucleotide primers - for
XX      PT      assaying DNA sample from respiratory secretion of a patient
XX      PS      infected with P carinii
XX      PS      Claim 2; Fig 3; 42pp; English.
XX      CC      This is the complete sequence of the insert of pAZ 112, with the
XX      CC      oligonucleotide primers indicated. The primers can be used in an
XX      CC      assay method for P. carinii. The assay is sensitive and can detect
XX      CC      P. carinii DNA in patients not yet showing clinical symptoms of
XX      CC      infection. P. carinii is a prime cause of opportunistic pneumonia
XX      CC      in patients with AIDS or immunosuppressed on oncology and transplant
XX      CC      units. See also AAO20063-Q20065 and AAO21396-Q21399.
XX      DR      Sequence 6785 BP; 2156 A; 998 C; 1134 G; 2497 T; 0 other;
XX      alignment_scores:
XX      Quality: 62.00      Length: 41
XX      Ratio: 2.000      Gaps: 0
XX      Percent Similarity: 75.610      Percent Identity: 31.707

alignment_block:
US-09-712-768-4 x AAO20065 ..
Align seg 1/1 to: AAO20065 from: 1 to: 6785
2 LeuGluIleValTrrpThrIleValProValIleLeuValIlePheIleG1 18
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2579 ATGAATTTGTTGGACAGTCCAGCTTCTTACTATATAGCCATTTC 2628

```

```

18 YAlAPheserleuProValIleuPheLysGlnGlnGlnIupheProGluLys 35
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2629 TTTTCCAAAGTTTCAAATTAATGTTATTAATGATCAAGTATGATCCAT 2678
35 sp1IleValIleAsnValGlnGly 42
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2679 CCATAACATTTAAAGCATACGT 2701

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAIT67748
seq_documentation_block:
ID      AAT67748 standard; DNA; 465 BP.
XX
AC      AAT67748;
XX
DN      30-JUL-1997 (first entry)
XX
DE      H. pylori secreted or periplasmic protein ORF 21687842.aa.
XX
KW      Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW      binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW      duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
XX
OS      Helicobacter pylori.
XX
XX      Key      Location/Qualifiers
XX      FT      1..465
XX      CDS
XX      /tag= a
XX      /transl_except= (pos: 166..168, aa: Xaa)
XX      /transl_except= (pos: 169..171, aa: Xaa)
XX      /transl_except= (pos: 295..297, aa: Xaa)
XX      /transl_except= (pos: 304..307, aa: Xaa)
XX      /transl_except= (pos: 319..321, aa: Xaa)
XX      /transl_except= (pos: 340..342, aa: Xaa)
XX      /transl_except= (pos: 358..360, aa: Xaa)
XX      /transl_except= (pos: 376..378, aa: Xaa)
XX      /note= "Xaa = unknown"

XX      W09640893-A1.
XX      PD      19-DEC-1996.
XX      PF      06-JUN-1996; 96WO-US09122.
XX      PR      01-APR-1996; 96US-0630405.
XX      PR      07-JUN-1995; 95US-0487032.
XX      PA      (ASTR ) ASTRA AB.
XX      PI      Berglinth OT, Smith D, Mellgaard BL;
XX      WPI: 1997-052306/05.
XX      DR      P-PSDB; AAW20218.
XX      FT      Helicobacter pylori nucleic acid sequences and related
XX      FT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX      PT      infection, and to detect Helicobacter
XX      PS      Claim 23; Page -: 1481pp; English.
XX      CC      This sequence encodes a H. pylori secreted or periplasmic protein.
XX      CC      The protein may be used in a vaccine to prevent or treat H. pylori
XX      CC      infection or to identify H. pylori polypeptide binding compounds,
XX      CC      useful as potential H. pylori life cycle activators or inhibitors.
XX      CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
XX      CC      overlapping contigs generated by mechanically shearing the bacterial
XX      CC      DNA. The sequences were analysed by computer evaluation. To
XX      CC      identify likely H. pylori antigens for vaccine development, the amino
XX      CC      acid sequences predicted from various ORF were analysed for significant
XX      CC      homology to other known or exported membrane proteins. Having identified
XX      CC      isolated from H. pylori by PCR amplification for recombinant polypeptide

```

Thu Dec 27 08:21:51 2001

production, e.g. in *E. coli* hosts.  
 Note: This DNA sequence is not reproduced in the specification and has been derived from the related specification, WO9719098.

Sequence 465 BP; 139 A; 71 C; 108 G; 139 T; 8 other:

alignment\_scores:      Length:      30  
                     quality:      60.00  
                     Ratio:      2.727      Gaps:      0  
                     Percent Similarity:      73.333      Percent Identity:      40.000

alignment\_block:  
 US-09-712-768-4 x AAT67748 ..

Align seg 1/1 to: AAT67748 from: 1 to: 465

```

2 LeuGluIleValTrpThrIleValProValIleLeuValPheIleG1 18
:.....:||||| :.....:
1 ATGAAATGGTGTGCTTATGTAGCGTATGTTTGTCTTTTACG 50
18 YAlapheserLeuProValLeuPheLysGlnGlnIuphe 31
||||| :.....:
51 GGCTGTAGAGTTGCTGTGAGTTATCAACTCAGAAATTT 90

```

seq\_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT: AAT77428

seq\_documentation\_block:  
 ID AAT77428 standard; DNA: 465 BP.

AC AAT77428;  
 DT 11-AUG-1997 (first entry)

DE H. pylori secreted or periplasmic protein ORF 21687842.aa.  
 XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
 XX secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;  
 KM detection; antisense; inhibition; ds.  
 XX Helicobacter pylori.

OS Helicobacter pylori.  
 XX Key Location/Qualifiers  
 FT CDS 1..465  
 FT /\*tag= a  
 FT /transl\_except= (pos: 166..168, aa: Xaa)  
 FT /transl\_except= (pos: 169..171, aa: Xaa)  
 FT /transl\_except= (pos: 295..297, aa: Xaa)  
 FT /transl\_except= (pos: 304..306, aa: Xaa)  
 FT /transl\_except= (pos: 319..321, aa: Xaa)  
 FT /transl\_except= (pos: 340..342, aa: Xaa)  
 FT /transl\_except= (pos: 358..360, aa: Xaa)  
 FT /transl\_except= (pos: 376..378, aa: Xaa)  
 FT /note= "Xaa = unknown"

XX WO9719098-A1:  
 PN 29-MAY-1997.  
 XX 15-NOV-1996; 96MO-US18542.  
 XX 17-NOV-1995; 95US-0561469.  
 XX (ASTR ) ASTRA AB.  
 XX Smith DH;  
 XX WPI: 1997-298052/27.  
 XX P-PSDB; AAM24610.  
 XX Helicobacter pylori nucleic acid sequences and related proteins -  
 XX used for diagnostics and therapeutics

XX  
 XX  
 CC Claim 1; page 89; 1481pp; English.  
 XX This sequence encodes a H. pylori secreted or periplasmic protein and  
 CC Helicobacter pylori has been strongly linked to chronic gastritis and  
 CC duodenal ulcer disease. The nucleic acid sequences of the invention  
 CC are used to evaluate compounds, especially activators or inhibitors of  
 CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
 CC sequence. The nucleic acid sequences, and corresponding subjects against H.  
 CC sequence, are also useful for generating vaccines for immunising subjects in  
 CC also useful for detecting the presence of these sequences are  
 CC pylori or for use in nucleic acid sequences of these sequences are  
 CC a sample. Antisense expression of a gene from Helicobacter species, H.  
 CC used to inhibit expression of a gene from Helicobacter species, H.  
 CC pylori whole genomic DNA was isolated and recombined and ligated to unique  
 CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
 CC complementary to the BstXI-cut PMX vectors, while the overhang is not  
 CC self-complementary. Therefore the linkers will not concatenate  
 CC will the cut vector re-ligate itself easily. The linker-adaptor inserts  
 CC were ligated to each of the 20 PMX vectors to construct a series of  
 CC shotgun subclone libraries. The purified DNA samples were then  
 CC sequenced.  
 CC Note: The ORF/protein reference number for this sequence was obtained  
 CC from the related specification, WO9640893.  
 XX Sequence 465 BP; 139 A; 71 C; 108 G; 139 T; 8 other;

alignment\_scores:      Length:      30  
                     quality:      60.00  
                     Ratio:      2.727      Gaps:      0  
                     Percent Similarity:      73.333      Percent Identity:      40.000

alignment\_block:  
 US-09-712-768-4 x AAT77428 ..

Align seg 1/1 to: AAT77428 from: 1 to: 465

```

2 LeuGluIleValTrpThrIleValProValIleLeuValPheIleG1 18
:.....:||||| :.....:
1 ATGAAATGGTGTGCTTATGTAGCGTATGTTTGTCTTTTACG 50
18 YAlapheserLeuProValLeuPheLysGlnGlnIuphe 31
||||| :.....:
51 GGCTGTAGAGTTGCTGTGAGTTATCAACTCAGAAATTT 90

```

seq\_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT: AAX30456

seq\_documentation\_block:  
 ID AAX30456 standard; DNA: 504 BP.

AC AAX30456;  
 DT 08-JUN-1999 (first entry)  
 DE H. pylori secreted protein ORF 06ep11202.21687842.c3\_35.  
 XX Vaccine; probe; diagnostic; ORF; cell envelope protein;  
 XX secreted protein; cellular protein; ds.  
 XX Helicobacter pylori.  
 XX WO9818323-A1.  
 XX 07-MAY-1998.  
 XX 28-OCT-1997; 97WO-US19575.  
 XX 14-JUL-1997; 97US-0891928.  
 XX 28-OCT-1996; 96US-0739150.  
 XX 06-DEC-1996; 96US-0759739.  
 XX (ASTR ) ASTRA AB.

XX  
 XX  
 CC Claim 1; page 89; 1481pp; English.  
 XX This sequence encodes a H. pylori secreted or periplasmic protein and  
 CC Helicobacter pylori has been strongly linked to chronic gastritis and  
 CC duodenal ulcer disease. The nucleic acid sequences of the invention  
 CC are used to evaluate compounds, especially activators or inhibitors of  
 CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
 CC sequence. The nucleic acid sequences, and corresponding subjects against H.  
 CC sequence, are also useful for generating vaccines for immunising subjects in  
 CC also useful for detecting the presence of these sequences are  
 CC pylori or for use in nucleic acid sequences of these sequences are  
 CC a sample. Antisense expression of a gene from Helicobacter species, H.  
 CC used to inhibit expression of a gene from Helicobacter species, H.  
 CC pylori whole genomic DNA was isolated and recombined and ligated to unique  
 CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
 CC complementary to the BstXI-cut PMX vectors, while the overhang is not  
 CC self-complementary. Therefore the linkers will not concatenate  
 CC will the cut vector re-ligate itself easily. The linker-adaptor inserts  
 CC were ligated to each of the 20 PMX vectors to construct a series of  
 CC shotgun subclone libraries. The purified DNA samples were then  
 CC sequenced.  
 CC Note: The ORF/protein reference number for this sequence was obtained  
 CC from the related specification, WO9640893.  
 XX Sequence 465 BP; 139 A; 71 C; 108 G; 139 T; 8 other;



```

XX Alm RA, Smith D;
XX
XX
XX WPI: 1998-271811/24.
XX P-PSDB: AAY10989.
XX
XX Helicobacter pylori nucleic acids and proteins - used to develop
XX products for the detection, prevention and treatment of H. pylori
XX infections
XX
XX Claims 3, 4; Page 122; 279pp; English.
XX
XX Recombinant or substantially pure preparations of H. pylori polypeptides
XX are disclosed, together with the nucleic acids encoding them. In all,
XX 73 ORFs are shown. The proteins are variously cell envelope proteins,
XX secreted proteins or other cellular proteins. Vaccines containing the
XX nucleic acids or proteins are claimed, as are probes containing at least
XX 8 nucleotides from the nucleic acid sequences. The vaccines are useful
XX for treating or reducing the risk of H. pylori infections, and the
XX probes can be used diagnostically for detecting the presence of
XX Helicobacter in a sample. The products are also of use in screening
XX for compounds having the ability to interfere with the H. pylori life
XX cycle or to inhibit H. pylori infection.
XX
XX Sequence 504 BP; 158 A; 81 C; 116 G; 149 T; 0 other.
XX
XX
XX alignment_scores:
XX   Quality: 60.00   Length: 30
XX   Percent Similarity: 73.333   Gaps: 0
XX   alignment_block:
XX   US-09-712-768-4 x AAX30456   Percent Identity: 40.000
XX
XX Align seg 1/1 to: AAX30456 from: 1 to: 504
XX
XX 2 LeuGIuIleValTrpHrIleValProValIleLeuValPheIleGI 18
XX   ::::||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 1 ATGAATTTGGTACGCTTATTGTAGCGCTTGTGTTGTTGTTTAAAG 50
XX 18 YAlaPheSerLeuProValIleuPheLysGInGluPhe 31
XX   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 51 GCCTGTAGAGTTCCTGGAGCTTATCAACCTCAAGAAATTT 90
XX
XX seq_name: /SID58/gcgdata/geneseq/geneseqn/NA1997.DAT:AA768220
XX
XX seq_documentation_block:
XX ID AA768220 standard; DNA: 543 BP.
XX
XX AC AA768220;
XX
XX DT 21-JUL-1997 (first entry)
XX
XX DE H. pylori secreted or periplasmic protein ORF hp2a11858orf5.
XX
XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bacterium; life cycle; activator;
XX diagnosis; ds.
XX
XX KW Helicobacter pylori.
XX
XX OS
XX
XX FH Key
XX
XX FT CDS
XX   Location/Qualifiers
XX   ..543
XX   /*tag= a
XX   /note= "no stop codon given"
XX
XX PD W09640893-A1.
XX
XX PD 19-DEC-1996.
XX
XX PF 06-JUN-1996; 96MO-US09122.

```

```

XX
XX PR 01-APR-1996; 96US-0630405.
XX PR 07-JUN-1995; 95US-0487032.
XX
XX PA (ASTR ) ASTRA AB.
XX
XX PI Berglindh OT, Smith D, Mellgaard BL;
XX
XX DR WPI: 1997-052306/05.
XX DR P-PSDB: AAM20967.
XX
XX PT Helicobacter pylori nucleic acid sequences and related
XX polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX infection, and to detect Helicobacter
XX
XX PS Claim 23; Page 971; 1481pp; English.
XX
XX CC This sequence encodes a H. pylori secreted or periplasmic protein.
XX CC The protein may be used in a vaccine to prevent or treat H. pylori
XX CC infection or to identify H. pylori life cycle activators or inhibitors.
XX CC Useful as potential H. pylori life cycle activators or inhibitors.
XX CC The genomic sequence of H. pylori (ATCC 55679) was determined from
XX CC overlapping contigs generated by mechanically shearing the bacterial
XX CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
XX CC and the predicted coding regions defined by computer evaluation. To
XX CC identify likely H. pylori antigens for vaccine development, the amino
XX CC acid sequences predicted from various ORF were analysed for significant
XX CC homology to other known or exported membrane proteins. Having identified
XX CC and determined the sequences of interest, particular regions can be
XX CC isolated from H. pylori by PCR amplification for recombinant polypeptide
XX CC production, e.g. in E. coli hosts.
XX
XX SQ Sequence 543 BP; 158 A; 87 C; 127 G; 171 T; 0 other.
XX
XX
XX alignment_scores:
XX   Quality: 60.00   Length: 30
XX   Percent Similarity: 73.333   Gaps: 0
XX   alignment_block:
XX   US-09-712-768-4 x AAT68220   Percent Identity: 40.000
XX
XX Align seg 1/1 to: AAT68220 from: 1 to: 543
XX
XX 2 LeuGIuIleValTrpHrIleValProValIleLeuValPheIleGI 18
XX   ::::||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 22 ATGAATTTGGTACGCTTATTGTAGCGCTTGTGTTGTTTAAAG 71
XX 18 YAlaPheSerLeuProValIleuPheLysGInGluPhe 31
XX   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 72 GCCTGTAGAGTTCCTGGAGCTTATCAACCTCAAGAAATTT 111
XX
XX seq_name: /SID58/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH54139
XX
XX seq_documentation_block:
XX ID AAH54139 standard; DNA: 3335 BP.
XX
XX AC AAH54139;
XX
XX DT 03-SEP-2001 (first entry)
XX
XX DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3503.
XX
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis; ds.
XX
XX OS Staphylococcus epidermidis.
XX
XX FH
XX
XX FT W0200134809-A2.
XX
XX PD 17-MAY-2001.
XX

```







CC terminates 14 nucleotides upstream from the E2 initiation codon.  
CC The E2 and E3 genes have been cloned and can be used for the  
CC recombinant production of BCV polypeptides, using e.g. Spodoptera  
CC frugiperda Sf9 insect cells as host cells. Glycosylated and  
CC non-glycosylated recombinant E2 and E3 are useful as components of  
CC vaccines directed toward preventing BCV infection, or reducing the  
CC severity of BCV infection, in bovine populations.  
XX

SQ Sequence 1305 BP; 331 A; 222 C; 247 G; 505 T; 0 other;

alignment\_scores:

Quality:	59.00	Length:	36
Ratio:	2.565	Gaps:	2
Percent Similarity:	63.889	Percent Identity:	38.889

alignment\_block:

US-09-712-768-4 x AAT89388 ..

Align seg 1/1 to: AAT89388 from: 1 to: 1305

```
4  ILeValTrpThr...ILeValProValValILeVal..... 15
   :::::|||||  |||:::|||||  |||:::|||||
183 CTTATATGAGACCTTAATCCTGCGCTGATCTGTGTAATAATATCATCT 232
16  ..... PheILeGlyAlaPheSerLeuProValLeuPheL 27
233 AAGCTGGCACTCCATTTTATGAGGTTTCACCTTACCGATTATTATA 282
27  ysgInGln 29
   |||
283 TTACACAG 290
```



```

; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 529
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-111-529

```

```

alignment_scores:
  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

## alignment\_block:

```
US-09-712-768-4 x US-09-328-111-529/rev ..
```

```
Align seg 1/1 to reverse of: US-09-328-111-529 from: 1 to: 251
```

```

2 LeuGIuIleValTrpThrIleValProValValIleLeuValPheIleGI 18
  ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
133 ATGGAACCGCTGACATCTGACCTGCGCCGCGCATCATCTAGTCCTATCGC 84
  : 18 yAlaPheSerLeuProValLeuPheLysGlnGlnIlePheProGluGly 35
  : 83 CCGCCCATCCCTACGATCCTTTACATACAGACGAGCTCAACGATCCCT 34
  : 35 spIleValIleAsnValGluGly 42
  : 33 CCTTACCATCAATCAATTGGC 11

```

```
seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-09-385-982-410
```

## seq\_documentation\_block:

```

; Sequence 410, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-410

```

## alignment\_scores:

```

  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

## alignment\_block:

```
US-09-712-768-4 x US-09-385-982-410 ..
```

```
Align seg 1/1 to: US-09-385-982-410 from: 1 to: 353
```

```

2 LeuGIuIleValTrpThrIleValProValValIleLeuValPheIleGI 18
  ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
218 ATGAACCGCTGACATCTGACCTGCGCCGCGCATCATCTAGTCCTATCGC 267
  : 18 yAlaPheSerLeuProValLeuPheLysGlnGlnIlePheProGluGly 35
  : 268 CCGCCCATCCCTACGATCCTTTACATACAGACGAGCTCAACGATCCCT 317
  : 35 spIleValIleAsnValGluGly 42
  : 318 CCTTACCATCAATCAATTGGC 340

```

```
seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-08-219-842-2
```

## seq\_documentation\_block:

```

; Sequence 2, Application US/08219842
; Patent No. 5565323
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; TITLE OF INVENTION: for Alzheimer's Disease
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/219,842
; FILING DATE: 30-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-AG 9504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-219-842-2

```

## alignment\_scores:

```

  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

## alignment\_block:

```
US-09-712-768-4 x US-08-219-842-2 ..
```

```
Align seg 1/1 to: US-08-219-842-2 from: 1 to: 754
```

```

2 LeuGIuIleValTrpThrIleValProValValIleLeuValPheIleGI 18
  ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
251 ATGAACCGCTGACATCTGACCTGCGCCGCGCATCATCTAGTCCTATCGC 300
  : 18 yAlaPheSerLeuProValLeuPheLysGlnGlnIlePheProGluGly 35
  : 301 CCGCCCATCCCTACGATCCTTTACATACAGACGAGCTCAACGATCCCT 350

```



35 sp1leval1leasnvalglugly 42  
: : : : :  
351 CCTTACCATCAATCAATTGGC 373

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-451-096-2

seq\_documentation\_block:

Sequence 2, Application US/08451096  
Patent No. 5760205  
GENERAL INFORMATION:  
APPLICANT: Parker, W. D.  
APPLICANT: Herrstadt, Corinna  
TITLE OF INVENTION: Diagnostic and Therapeutic Compositions  
TITLE OF INVENTION: for Alzheimer's Disease  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,096  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/219,842  
FILING DATE: 30-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-AG 9504  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 754 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-451-096-2

alignment\_scores:

Quality: 58.00 Length: 41  
Ratio: 2.000 Gaps: 0  
Percent Similarity: 70.732 Percent Identity: 36.585

alignment\_block:

US-09-712-768-4 x US-08-451-096-2 ..

Align seg 1/1 to: US-08-451-096-2 from: 1 to: 754

2 Leugluileval1trphtlleval1proval1alleuval1phei1eg1 18  
: : : : :  
251 ATAGAAACGCTGTAAGTATCTGCGCCGCGCATCATCTAGTCTCATCGC 300  
18 yalapheserleuproval1leuphelysglncglupheproglugly 35  
: : : : :  
301 CCTCCATCCCTACGATCCTTACATACAGACGAGGTCAACGATCCCT 350  
35 sp1leval1leasnvalglugly 42  
: : : : :  
351 CCTTACCATCAATCAATTGGC 373

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-09-385-982-539

seq\_documentation\_block:

Sequence 539, Application US/09385982  
Patent No. 6262334  
GENERAL INFORMATION:  
APPLICANT: ENDEGE, WILSON O., ET AL.  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
TITLE OF INVENTION: PRODUCTS II  
FILE REFERENCE: CCDNA-260XX  
CURRENT APPLICATION NUMBER: US/09/385,982  
CURRENT FILING DATE: 1999-08-30  
EARLIER APPLICATION NUMBER: 09/328,111  
EARLIER FILING DATE: 1999-06-08  
EARLIER APPLICATION NUMBER: 60/117,393  
EARLIER FILING DATE: 1999-01-27  
EARLIER APPLICATION NUMBER: 60/098,639  
NUMBER OF SEQ ID NOS: 544  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 539  
LENGTH: 789  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(789)  
OTHER INFORMATION: n = A,T,C or G  
US-09-385-982-539

alignment\_scores:

Quality: 58.00 Length: 41  
Ratio: 2.000 Gaps: 0  
Percent Similarity: 70.732 Percent Identity: 36.585

alignment\_block:

US-09-712-768-4 x US-09-385-982-539/rev ..

Align seg 1/1 to reverse of: US-09-385-982-539 from: 1 to: 789

2 Leugluileval1trphtlleval1proval1alleuval1phei1eg1 18  
: : : : :  
133 ATAGAAACGCTGTAAGTATCTGCGCCGCGCATCATCTAGTCTCATCGC 84  
18 yalapheserleuproval1leuphelysglncglupheproglugly 35  
: : : : :  
83 CCTCCATCCCTACGATCCTTACATACAGACGAGGTCAACGATCCCT 34  
35 sp1leval1leasnvalglugly 42  
: : : : :  
33 CCTTACCATCAATCAATTGGC 11  
seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-413-740A-2  
seq\_documentation\_block:  
Sequence 2, Application US/08413740A  
Patent No. 6171859  
GENERAL INFORMATION:  
APPLICANT: HERRNSTADT, CORINNA  
APPLICANT: PARKER, WILLIAM D.  
APPLICANT: DAVIS, ROBERT  
APPLICANT: MILLER, SCOTT W.  
TITLE OF INVENTION: Diagnosis, Therapy and Cellular and  
TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial  
NUMBER OF SEQUENCES: 206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA

```

: ZIP: 20036-5405
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/413,740A
: FILING DATE:
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04063
: FILING DATE: 30-MAR-1995
: APPLICATION NUMBER: 08/413,740
: FILING DATE: 30-MAR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Bonham, David B.
: REGISTRATION NUMBER: 34297
: REFERENCE/DOCKET NUMBER: 2105/7
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 429-1776
: TELEFAX: (202) 429-0796
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-413-740A-2

```

```

alignment_scores:
  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

```

alignment_block:
  US-09-712-768-4 x US-08-413-740A-2

```

```

Align seg 1/1 to: US-08-413-740A-2 from: 1 to: 854

```

```

2 LeuGluIleValTrpThrIleValProValIleLeuValPheIleG1 18
: : : : : | | | | | : : : : : : : : : : : : : : : :
251 ATGAGAACCGTGTGACTATCTGCGCCGATCATCTAGTCTCATCGC 300
18 yAlaPheSerLeuProValIleuPheLysGlnGlnIupheProGluG1A 35
: : : : : | | | | | : : : : : : : : : : : : : : : :
301 CCTCCCATCCCTACGATCTCTTACATTAACAGACGAGGTCAACGATCCCT 350
35 splIleValIleAsnValGluGly 42
: : : : : | | | | | : : : : : : : : : : : : : : : :
351 CCTTACCATCAATCAATTGGC 373

```

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-04063-2

```

```

seq_documentation_block:
: Sequence 2, Application PC/TUS9504063
: GENERAL INFORMATION:
: APPLICANT: HERRNSTADT, CORINNA
: APPLICANT: PARKER, WILLIAM D.
: APPLICANT: DAVIS, ROBERT
: APPLICANT: MILLER, SCOTT W.
: TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
: TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
: NUMBER OF SEQUENCES: 206
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenyon & Kenyon
: STREET: 1025 Connecticut Avenue, N.W.

```

```

: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20036-5405
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04063
: FILING DATE: 30-MAR-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bonham, David B.
: REGISTRATION NUMBER: 34297
: REFERENCE/DOCKET NUMBER: 2105/7
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 429-1776
: TELEFAX: (202) 429-0796
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-04063-2

```

```

alignment_scores:
  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

```

alignment_block:
  US-09-712-768-4 x PCT-US95-04063-2

```

```

Align seg 1/1 to: PCT-US95-04063-2 from: 1 to: 854

```

```

2 LeuGluIleValTrpThrIleValProValIleLeuValPheIleG1 18
: : : : : | | | | | : : : : : : : : : : : : : : : :
251 ATGAGAACCGTGTGACTATCTGCGCCGATCATCTAGTCTCATCGC 300
18 yAlaPheSerLeuProValIleuPheLysGlnGlnIupheProGluG1A 35
: : : : : | | | | | : : : : : : : : : : : : : : : :
301 CCTCCCATCCCTACGATCTCTTACATTAACAGACGAGGTCAACGATCCCT 350
35 splIleValIleAsnValGluGly 42
: : : : : | | | | | : : : : : : : : : : : : : : : :
351 CCTTACCATCAATCAATTGGC 373

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-415-751-27

```

```

seq_documentation_block:
: Sequence 27, Application US/08415751
: Patent No. 5643772
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUT, JIRI
: TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
: TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
: TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
: TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
: TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PHILLIPS, MOORE, LEMPIO & FINLEY

```

```
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071,880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/891,301
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolezalova
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.19-2 (HMD)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
US-08-415-751-27

alignment_scores:
Quality: 58.00 Length: 23
Ratio: 3.053 Gaps: 0
Percent Similarity: 82.609 Percent Identity: 34.783

alignment_block:
US-09-712-768-4 x US-08-415-751-27/rev ..

Align seg 1/1 to reverse of: US-08-415-751-27 from: 1 to: 1086

3 GuilevaltrpThrllevalprovalvalleleuvalpheileglyal 19
||||:||||| :||||| :||||| :|||
361 GAGTGTGTGCTGTTGCTGCTTAGATAGTGTAGTACTAGTGTGTT 312
19 apheSerleuprovalleu 25
:|||||:|||||:
311 GCTTGACCTTCAGTTGTA 293

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-415-751-28

seq_documentation_block:
; Sequence 28, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; VECTOR OF INVENTION: AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; NUMBER OF SEQUENCES: 50
```

```
CORRESPONDENCE ADDRESS:
ADDRESSEE: PHILLIPS, MOORE, LEMPJO & FINLEY
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071,880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/891,301
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolezalova
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.19-2 (HMD)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
US-08-415-751-28

alignment_scores:
Quality: 58.00 Length: 23
Ratio: 3.053 Gaps: 0
Percent Similarity: 82.609 Percent Identity: 34.783

alignment_block:
US-09-712-768-4 x US-08-415-751-28 ..

Align seg 1/1 to: US-08-415-751-28 from: 1 to: 1086

3 GuilevaltrpThrllevalprovalvalleleuvalpheileglyal 19
||||:||||| :||||| :||||| :|||
726 GAGTGTGTGCTGTTGCTGCTTAGATAGTGTAGTACTAGTGTGTT 775
19 apheSerleuprovalleu 25
:|||||:|||||:
776 GCTTGACCTTCAGTTGTA 794

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-097-889-2

seq_documentation_block:
; Sequence 2, Application US/09097889
; Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: HERNSTADT, CORINA
; APPLICANT: GHOSH, SOUMITRA S.
; APPLICANT: DAVIS, ROBERT E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; TITLE OF INVENTION: EXTRA MITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
```

OFFENDING SYSTEM: IBM P.C. DOS (version 3.30)

FILING DATE: 02-MAY-94

FILING DATE: 02-MAY-94

1 APPLICATION NUMBER: 08/210,250  
2 FILING DATE: 15-MAR-94  
3 PRIOR APPLICATION DATA:  
4 APPLICATION NUMBER: 07/803,842  
5 FILING DATE: 09-DEC-91  
6 PRIOR APPLICATION DATA:  
7 APPLICATION NUMBER: 07/726,606  
8 FILING DATE: 09-JUL-91  
9 PRIOR APPLICATION DATA:  
10 APPLICATION NUMBER: 07/685,101  
11 FILING DATE: 10-APR-91  
12 ATTORNEY/AGENT INFORMATION:  
13 NAME: Fasse, J. Peter  
14 REGISTRATION NUMBER: 32,983  
15 REFERENCE/DOCKET NUMBER: 04766/015WOL  
16 TELECOMMUNICATION INFORMATION:  
17 TELEPHONE: (617) 542-5070  
18 TELEFAX: (617) 542-8906  
19 TELEX: 200154  
20 INFORMATION FOR SEQ ID NO: 4:  
21 SEQUENCE CHARACTERISTICS:  
22 LENGTH: 1106  
23 TYPE: nucleic acid  
24 STRANDEDNESS: single  
25 TOPOLOGY: linear  
26 CACT-UST95-03032-4

```

NAME: Scott, Watson T
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: WTS/5683/91535/WMH
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 cush
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1510 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-759-568-4

alignment_scores:
Quality: 56.00 Length: 28
Ratio: 2.667 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 42.857

alignment_block:
US-09-712-768-4 x US-07-759-568-4 ..

Align seg 1/1 to: US-07-759-568-4 from: 1 to: 1510

```

```

; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1748 bases
; TYPE: nucleic acid
; STRADEDNESS: single
; TOPOLOGY: linear
;
US-08-202-056-8

```

alignment_scores:		
Quality:	56.00	Length: 28
Ratio:	2.667	Gaps: 1
Percent Similarity:	75.000	Percent Identity: 42.857

```
alignment_block:
US-09-712-768-4 x US-08-202-056-8 . .
```

Align seg 1/1 to: US-08-202-056-8 from: 1 to: 1748

17 lecltylaalpheserleauprovalleaupehyis 27  
 740 GCTACGGATTACCCCTGGCTACCTGCTTTTAAAG 771



LOCUS BE518487 705 bp mRNA EST 21-MAY-2001  
 DEFINITION EST00063 Atlantic salmon Lambda Zap Express testis cDNA library  
 Salmu salar cDNA clone TSM-119 5' similar to cytochrome C oxidase  
 chain II, mRNA sequence.  
 ACCESSION BE518487  
 VERSION BE518487.1 GI:12621507  
 KEYWORDS EST.  
 SOURCE Atlantic salmon.  
 ORGANISM Salmu salar  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
 REFERENCE 1 (bases 1 to 705)  
 DAVEY, G.C., CAPLICE, N.C., MARTIN, S.A. and POWELL, R.  
 TITLE A survey of genes in the Atlantic salmon (Salmu salar) as  
 identified by expressed sequence tags  
 JOURNAL Gene 263 (1-2), 121-130 (2001)  
 MEDLINE 21125223  
 COMMENT Contact: Davey GC  
 Department of Microbiology  
 National University of Ireland Galway (NUIG)  
 Galway, Ireland  
 Tel: 00353 91 524411 2254  
 Fax: 00353 91 525700  
 Email: grace.davey@nuigalway.ie  
 Insert Length: 719 Std Error: 0.00  
 Plate: Testis S+M (E2) row: f column: 08  
 Seq primer: M13 reverse primer = caggaacagctatgacc  
 High quality sequence stop: 719  
 POLYA=yes.

FEATURES  
 source location/Qualifiers  
 1..705  
 /organism="Salmu salar"  
 /db\_xref="taxon:8030"  
 /clone="TSM-119"  
 /clone\_lib="Atlantic salmon lambda zap Express testis cDNA  
 library"  
 /sex="male"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /note="Organ: testis; Vector: Lambda ZAP Express; Site\_1:  
 EcoRI; Site\_2: XhoI; An Atlantic salmon testis cDNA  
 library was constructed using the Lambda zap  
 express/gigapack cloning kit (Stratagene cloning systems).  
 cDNA synthesis was carried out using an oligo-(dT) primer  
 for the reverse transcription of 5ug of mRNA and the  
 library was constructed by directional cloning EcoRI-XhoI  
 based on manufacturers instructions. An insert:vector  
 ligation ratio of 1:5 was chosen as most optimum. The  
 lambda library was packaged with gigapack III gold  
 packaging extracts and plated on the E. coli cell line  
 XL1-blue MRF".

BASE COUNT 205 a 190 c 116 g 192 t 2 others  
 ORIGIN

alignment\_scores: Quality: 79.00 Length: 41  
 Ratio: 2.469 Gaps: 0  
 Percent Similarity: 78.049 Percent Identity: 39.024

alignment\_block:  
 US-09-712-768-4 x BE518487 ..  
 Align seg 1/1 to: BE518487 from: 1 to: 705

```

2 Leugluilevaltrphtlilevalprovalvalilleuvalpheiieci 18
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 ATCGAATCTTTGAGCTGCTCCAGCAGTATCTCATCTTATTCG 228
18 yalapheserleuprovalleuphelsglinglupheprogluglya 35
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 CTTCCCTCTTCAATCTTACCTTATGACGAAATTATGACCAC 278

```

35 splevalilleasvalgluglya 42  
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 279 ACCTTACTTATTAAGCAATGGCT 301

seq\_name: gb\_estl:AW600763

seq\_documentation\_block:  
 LOCUS AW600763 661 bp mRNA EST 30-OCT-2000  
 DEFINITION ESTPM2293 Penaeus monodon's total hemocyte cDNA library (#2)  
 Penaeus monodon cDNA clone PMC293 5', mRNA sequence.  
 ACCESSION AW600763  
 VERSION AW600763.1 GI:11037892  
 KEYWORDS EST.  
 SOURCE Black tiger shrimp.  
 ORGANISM Penaeus monodon  
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;  
 Penaeidae; Penaeus.  
 REFERENCE 1 (bases 1 to 661)  
 WONGSANTICHON, J., BOONCHOY, C., UDOMKIT, A., PANYIM, S. and  
 SONTAYANON, B.  
 TITLE Expressed Sequence Tags from Black Tiger Prawn's Hemocytes  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Sontayanon, B.  
 Prawn Molecular Biology  
 Institute of Molecular Biology and Genetics, Mahidol University,  
 Salaya campus  
 Phuthamonthon 4 Rd., Phuthamonthon District, Nakhon Pathom, 73170  
 Thailand  
 Tel: 66 2 4419003  
 Fax: 66 2 4419006  
 Email: scbst@mahidol.ac.th  
 Similar to g14582668|emb|X84350.2|MTPNID1 Penaeus notialis  
 mitochondrial genome, partial (Acc.no.4582668) cytochrome c oxidase  
 subunit II Poly A tract is found after the end of this sequence.  
 Seq primer: T3  
 POLYA=yes.

FEATURES  
 source location/Qualifiers  
 1..661  
 /organism="Penaeus monodon"  
 /db\_xref="taxon:6687"  
 /clone="PMC293"  
 /clone\_lib="Penaeus monodon's total hemocyte cDNA library  
 (#2)"  
 /tissue\_type="hemocytes"  
 /dev\_stage="16-17 g average body weight (farm-raised)"  
 /lab\_host="E.coli XL1-Blue MRF"  
 /note="Vector: lambdaZAP II; Site\_1: EcoR I; Site\_2: Xho  
 I; The library was prepared using protocol given by  
 supplier, Stratagene, Inc."

BASE COUNT 209 a 106 c 97 g 248 t 1 others  
 ORIGIN

alignment\_scores: Quality: 77.00 Length: 41  
 Ratio: 2.406 Gaps: 0  
 Percent Similarity: 78.049 Percent Identity: 36.585

alignment\_block:  
 US-09-712-768-4 x AW600763 ..  
 Align seg 1/1 to: AW600763 from: 1 to: 661

```

2 Leugluilevaltrphtlilevalprovalvalilleuvalpheiieci 18
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 ATGGAATTTGATGAGCACTATCCCTCAATTAATTTTATTCG 209
18 yalapheserleuprovalleuphelsglinglupheprogluglya 35
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 ACTTCTCTTTCAGATTACTTATCTTATGATGAGTAATTAATATCCAA 259

```





```

seq_documentation_block:
LOCUS       AV670026             467 bp             mRNA                      EST             22-SEP-2000
DEFINITION   AV670026 OLHNI cell line cDNA library (OLb) Oryzias latipes cDNA
              clone OLb16.03d similar to cytochrome-c oxidase (EC 1.9.3.1) chain
              II (Atlantic salmon mitochondrion), mRNA sequence.
ACCESSION    AV670026
VERSION      AV670026.1  GI:9935824
KEYWORDS     EST.
SOURCE       Japanese medaka.
ORGANISM     Oryzias latipes
REFERENCE    Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
              Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei
              Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
              Belontiiformes: Adrianchthyidae; Oryziinae; Oryzias.
              1 (bases 1 to 467)
AUTHORS      Naruse,K., Tanaka,M., Shima,A. and Mitani,H.
TITLE        Medaka EST project in University of Tokyo
JOURNAL      Unpublished (2000)
COMMENT      Contact: Kiyoshi Naruse
              Department of Biological Sciences
              Graduate School of Science, University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
              Tel: 81-3-5841-4443
              Fax: 81-3-5841-4410
              Email: naruse@biol.s.u-tokyo.ac.jp
              This clone was isolated from OLHNI cell line cDNA library (OLb) 5
              end sequences.
FEATURES
     source                Location/Qualifiers
         1..467
            /organism="Oryzias latipes"
            /strain="HN1"
            /db_xref="taxon:80990"
            /clone="OLb16.03d"
            /clone_lib="OLHNI cell line cDNA library (OLb)"
BASE COUNT      138 a 119 c 77 g 133 t
ORIGIN
alignment_scores:
      Quality: 75.00      Length: 41
      Ratio: 2.344      Gaps: 0
      Percent Similarity: 78.049      Percent Identity: 34.146
alignment_block:
US-09-712-768-4 x AV670026 ..
Align seg 1/1 to: AV670026 from: 1 to: 467
2 leugluilevalttrprrlllevalprovalvallelevalpheilegl 18
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
136 attgaaattatcttgacccttgcttcgacgaattatcttatcttgatcgc 185
18 yalapheserleuprovalleuphelsgningluphproglucly 35
: ||||| :|||::: :|||::: :|||::: :|||::: :|||:::
186 cttaccctcccttgccattctttaccttatagatgaattatgaccccc 235
235 spllevalilleasnvalglucly 42
:::|||||:::|||||
236 acctgacacattaaagccatagacc 258
seq_name: gb_est1:AV669351
seq_documentation_block:
LOCUS       AV669351             469 bp             mRNA                      EST             22-SEP-2000
DEFINITION   AV669351 OLHNI cell line cDNA library (OLb) Oryzias latipes cDNA
              clone OLb05.06h similar to cytochrome-c oxidase (EC 1.9.3.1) chain
              II (Atlantic salmon mitochondrion), mRNA sequence.
ACCESSION    AV669351
VERSION      AV669351.1  GI:9934188
KEYWORDS     EST.
SOURCE       Japanese medaka.
ORGANISM     Oryzias latipes

```

Thu Dec 27 08:21:57 2001

us-09-712-768-4.1st

Page 5

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:  
Acanthomorpha: Acanthopterygii: Percomorpha: Atherinomorpha:  
Belontiiformes: Adrianchthyidae: Oryziinae: Oryzias.  
1 (bases 1 to 469) Shima, A. and Mitani, H.  
Narusue, K., Tanaka, M.,  
Medaka EST Project in University of Tokyo  
Unpublished (2000)  
Contact: Kiyoshi Narusue  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo  
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4410  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'  
end sequences.  
Location/Qualifiers  
1..469  
/organism="Oryzias latipes"  
/strain="HN1"  
/db\_xref="taxon:8090"  
/clone="OLb05.06h"  
/clone\_lib="OLHNI cell line cDNA library (OLb) 5'  
others  
136 a 126 c 78 g 127 t  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 75.00 Length: 41  
Ratio: 2.344 Gaps: 0  
Percent Similarity: 78.049 Percent Identity: 34.146  
Percent Identity: 78.049  
alignments\_block:  
US-09-712-768-4 x AV669351 ..  
Align seg 1/1 to: AV669351 from: 1 to: 469  
2 LeuGluIleValTrpThrIleValProValValIleLeuValPheIleG1 18  
18 ATGGAATTAATCTGACCTGCTCCAGCAGCATTAATCTTAATCTGATCGC 136  
18 ATGGAATTAATCTGACCTGCTCCAGCAGCATTAATCTTAATCTGATCGC 136  
18 yAlaPheSerLeuProValLeuPheLysGlnGlnGlnPheProGluGly 35  
137 CTACCTCCCTCCGATCTTCTTACCTTAATGATGAATTAATGACCCCC 186  
35 spIleValIleAsnValGluGly 42  
187 ACCTGACAAATTAAGCCATGAGC 209  
seq\_name: gb\_est1:AV670767  
seq\_documentation\_block: 486 bp mRNA EST 22-SEP-2000  
LOCUS AV670767 OLHNI cell line cDNA library (OLb) Oryzias latipes cDNA  
DEFINITION AV670767.05f similar to cytochrome-c oxidase (EC 1.9.3.1) chain  
II (Atlantic salmon mitochondrion), mRNA sequence.  
ACCESSION AV670767  
VERSION AV670767.1 GI:9936565  
KEYWORDS EST.  
SOURCE Japanese medaka.  
ORGANISM Oryzias latipes  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:  
Acanthomorpha: Acanthopterygii: Percomorpha: Atherinomorpha:  
Belontiiformes: Adrianchthyidae: Oryziinae: Oryzias.  
1 (bases 1 to 486) Shima, A. and Mitani, H.  
Narusue, K., Tanaka, M.,  
Medaka EST Project in University of Tokyo  
Unpublished (2000)  
Contact: Kiyoshi Narusue  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4443  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'  
end sequences.  
Location/Qualifiers  
1..486  
/organism="Oryzias latipes"  
/strain="HN1"  
/db\_xref="taxon:8090"  
/clone="OLb27.05f"  
/clone\_lib="OLHNI cell line cDNA library (OLb) 5'  
136 t  
141 a 128 c 81 g  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 75.00 Length: 41  
Ratio: 2.344 Gaps: 0  
Percent Similarity: 78.049 Percent Identity: 34.146  
Percent Identity: 78.049  
alignments\_block:  
US-09-712-768-4 x AV670767 ..  
Align seg 1/1 to: AV670767 from: 1 to: 486  
2 LeuGluIleValTrpThrIleValProValValIleLeuValPheIleG1 18  
18 ATGGAATTAATCTGACCTGCTCCAGCAGCATTAATCTTAATCTGATCGC 173  
124 ATGGAATTAATCTGACCTGCTCCAGCAGCATTAATCTTAATCTGATCGC 173  
18 yAlaPheSerLeuProValLeuPheLysGlnGlnGlnPheProGluGly 35  
174 CTACCTCCCTCCGATCTTCTTACCTTAATGATGAATTAATGACCCCC 223  
35 spIleValIleAsnValGluGly 42  
224 ACCTGACAAATTAAGCCATGAGC 246  
seq\_name: gb\_est1:AV670936  
seq\_documentation\_block: 502 bp mRNA EST 22-SEP-2000  
LOCUS AV670936 OLHNI cell line cDNA library (OLb) Oryzias latipes cDNA  
DEFINITION AV670936.07a similar to cytochrome-c oxidase (EC 1.9.3.1) chain  
II (Atlantic salmon mitochondrion), mRNA sequence.  
ACCESSION AV670936  
VERSION AV670936.1 GI:9936734  
KEYWORDS EST.  
SOURCE Japanese medaka.  
ORGANISM Oryzias latipes  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:  
Acanthomorpha: Acanthopterygii: Percomorpha: Atherinomorpha:  
Belontiiformes: Adrianchthyidae: Oryziinae: Oryzias.  
1 (bases 1 to 502) Shima, A. and Mitani, H.  
Narusue, K., Tanaka, M.,  
Medaka EST Project in University of Tokyo  
Unpublished (2000)  
Contact: Kiyoshi Narusue  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo  
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4410  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'  
end sequences.  
Location/Qualifiers  
1..502  
/organism="Oryzias latipes"  
/strain="HN1"  
/db\_xref="taxon:8090"



Penaeidae; Marsupenaeus.

AUTHORS Rojtimakorn,J., Hirono,I., Aoki,T., Itami,T. and Takahashi,Y.  
 TITLE Gene expression in non-infected and viral infected hemocytes of  
 kuruma prawn (Penaeus japonicus)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Takashi Aoki  
 Aquatic Biosciences  
 Tokyo University of Fisheries  
 Konan 4-5-7, Tokyo, Minato-ku 108-8477, Japan  
 Tel: 81-3-5463-0689  
 Fax: 81-3-5463-0690  
 Email: aoki@tokyo-u-fish.ac.jp.

FEATURES  
 source  
 1 .579  
 location/Qualifiers  
 /organism="Marsupenaeus japonicus"  
 /db\_xref="taxon:27405"  
 /clone="PJH325"  
 /clone\_id="Marsupenaeus japonicus adult"  
 /cell\_type="hemocytes"  
 /dev\_stage="adult"  
 /note="Vector: lambda ZAP II"

BASE COUNT 202 a 95 c 121 g 160 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 75.00 Length: 41  
 Ratio: 2.419 Gaps: 0  
 Percent Similarity: 75.610 Percent Identity: 36.585

alignment\_block:  
 US-09-712-768-4 x AU175272/rev ..

Align seg 1/1 to reverse of: AU175272 from: 1 to: 579

```

2 LeuGluIleValTrpThrIleValProValValIleLeuValPheIleG1 18
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
479 ATTGAATGTGATGGACACTCTCGGCCCTTATCTTATCTCATCTCGC 430
18 yAlaPheSerLeuProValLeuPheLysGlnGlnGluPheProGluG1YA 35
|||||:::|||||:::|||||:::|||||:::|||||:::
429 CCTACTCTCTTACGATTTACTTATTACTAGATGAGATTCAATATCCNA 380
35 spIleValIleAsnValGluG1Y 42
:::|||||:::|||||:::|||||:::|||||:::|||||:::
379 GAGTTACATTGAGACACATTCGT 357
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:01:48 ; Search time 39.94 seconds  
(without alignments)  
81.603 Million cell updates/sec

Title: US-09-712-768-4

Perfect score: 222  
Sequence: 1 PLEIWTIVPVILVIFGAF.....LFKQEPPEGDIVNEGRS 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A-Geneseq\_1101.\*  
2: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
23: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	222	100.0	44	22	AAV97751
2	75	33.8	129	21	AA839150
3	62	27.9	243	13	AA821411
4	60	27.0	154	18	AAW20218
5	60	27.0	154	18	AAW24610
6	60	27.0	167	19	AAV10989
7	59	27.0	181	18	AAW20967
8	59	26.6	20	22	AAV97755
9	59	26.6	55	21	AA803705
10	59	26.6	197	22	AA890866
11	56.5	25.5	593	22	AA805103

12	56	25.2	359	22	AA892151
13	56	25.2	359	22	AA879573
14	55	24.8	132	21	AA839151
15	54	24.3	1019	19	AAW71527
16	54	24.3	1020	19	AAW98661
17	54	24.3	1020	22	AAW46344
18	53.5	24.1	363	18	AAW08437
19	53.5	24.1	363	18	AAW08437
20	53	23.9	83	22	AAW24378
21	53	23.9	391	21	AAW06185
22	53	23.9	391	21	AAW49300
23	52.5	23.6	113	21	AAW87356
24	52.5	23.6	114	19	AAW37817
25	52.5	23.6	314	20	AAW88528
26	52	23.4	40	22	AAW75937
27	52	23.4	40	22	AAW75938
28	52	23.4	1291	16	AAW75201
29	52	23.4	1816	21	AAW95440
30	51.5	23.2	259	20	AAW35625
31	51.5	23.2	691	11	AAW05234
32	51.5	23.2	1169	16	AAW76544
33	51	23.0	193	21	AAW37400
34	51	23.0	193	21	AAW37422
35	51	23.0	212	21	AAW37398
36	51	23.0	212	21	AAW37420
37	51	23.0	385	21	AAW06186
38	51	23.0	385	21	AAW49301
39	51	23.0	407	19	AAW80618
40	51	23.0	407	21	AAW81718
41	51	23.0	407	22	AAW01042
42	51	23.0	437	21	AAW13750
43	51	23.0	437	21	AAW20978
44	51	23.0	438	21	AAW38871
45	51	23.0	447	21	AAW13749

# ALIGNMENTS

RESULT 1	
ID	AAV97751 standard; Protein: 44 AA.
XX	AAV97751;
AC	06-AUG-2001 (first entry)
XX	
DE	G. oxydans cytochrome C oxidase (COII) protein sequence fragment.
XX	
KW	Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;
KW	oxidative fermentation; electron transfer; respiratory chain; L-sorbose;
KW	2-Keto-L-gluconic acid; 2KG production; aldehyde production;
KW	carboxylic acid production; ketone production.
XX	
OS	Glucanobacter oxydans.
XX	
FN	EP1103603-A2.
XX	
PD	30-MAY-2001.
XX	
PR	14-NOV-2000; 2000EP-0124785.
XX	
PR	17-NOV-1999; 99EP-0122842.
XX	
PA	(HOFF) HOFFMANN LA ROCHE & CO AG F.
XX	
PI	Asakura A, Hoshino T, Shinjoh M;
XX	
DR	WPI; 2001-357953/38.
XX	
XX	N-PSDB; AAA91490.
PT	New cytochrome C oxidase complex having cytochrome C oxidase activity from Glucanobacter oxydans DSM 4025, useful in mediating electron

PT transfer in respiratory chain or producing 2-keto-L-gulononic acid from  
 PT L-sorbose or D-sorbitol  
 XX  
 PS Claim 6; Page 24; 42pp; English.

CC This sequence represents a fragment of the gluconobacter oxydans  
 CC cytochrome C oxidase, COI1. The invention relates to a cytochrome c  
 CC oxidase COI complex. The COI complex is useful in improving oxidative  
 CC fermentation and is an essential component mediating electron transfer  
 CC in the respiratory chain. The recombinant microorganism and the  
 CC cytochrome c oxidase may be used in the genetic preparation of a  
 CC recombinant COI complex and in the production of 2-keto-L-gulononic acid  
 CC (2KGA) from L-sorbose or D-sorbitol in a culture medium. The COI is also  
 CC useful as a terminal oxidase, oxidising cytochrome c, an electron  
 CC acceptor from an enzyme belonging to dehydrogenase c, for the production of  
 CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,  
 CC especially the production of 2KGA from L-sorbose or D-sorbitol.

XX Sequence 44 AA;

Query Match 100.0%; Score 222; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLEIWTIVPVVILVFIAGSLPVLFKQEPFEGDIVINVEGRS 44  
 Db 1 pleiwtivpvvillvfiagslpvlfkqepfegdivinvegrs 44

RESULT 2

AAB39150  
 ID AAB39150 standard; Protein; 129 AA.

XX AAB39150;

DT 02-FEB-2001 (first entry)

DE Human secreted protein BLAST search protein SEQ ID NO: 117.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

OS WO200058513-A1.

XX 05-OCT-2000.

PD 22-MAR-2000; 2000WO-US07506.

XX 26-MAR-1999; 99US-0126505.

PR 17-DEC-1999; 99US-0172412.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-594649/56.

XX Forty-nine polynucleotide sequences, and their encoded secreted  
 PT polypeptides, used in the treatment and diagnosis of cancers,  
 PT autoimmune disorders, and skin disorders -

PS Disclosure; Page 394; 413pp; English.

XX The invention relates to the isolation of genes AAC73865-C73913 encoding  
 CC the human secreted proteins AAB39093-839141. This sequence represents a  
 CC peptide fragment homologous to the protein encoded by the gene isolated

CC in the present invention. The sequence is a search result from a BLASTX  
 CC homology search. The genes and proteins are useful for preventing,  
 CC ameliorating or treating medical conditions, e.g. by protein or gene  
 CC therapy. The genes are isolated from a range of human tissues disclosed  
 CC in the specification. The nucleic acids, proteins, antibodies and  
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such  
 CC as myocardial ischaemias; (d) wound healing; (e) neurological diseases  
 CC as viral, bacterial, fungal and parasitic infections.

XX Sequence 129 AA;

Query Match 33.8%; Score 75; DB 21; Length 129;  
 Best Local Similarity 39.0%; Pred. No. 0.0022;  
 Matches 16; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

QY 2 LEIWTIVPVVILVFIAGSLPVLFKQEPFEGDIVINVEG 42  
 Db 36 metwtllpallvllvllpallrllymtdevrnpstlksig 76

RESULT 3

AAR21411  
 ID AAR21411 standard; Protein; 243 AA.

XX AAR21411;

DT 01-APR-1992 (first entry)

DE Cytochrome oxidase subunit II.

XX Pneumonia; assay; AIDS; immunosuppressed.

OS Pneumocystis carinii.

XX WO9119005-A.

XX 12-DEC-1991.

XX 31-MAY-1991; 91WO-GB00869.

PR 01-JUN-1990; 90GB-0012196.

XX (ISIS-) ISIS INNOVATION LTD.

PI Wakefield AE, Hopkin JM, Moxon ER;

XX WPI; 1992-007487/01.

DR N-PSDB; AAO20065.

XX New DNA sequences which act as oligo:nucleotide primers - for  
 PT assaying DNA sample from respiratory secretion of a patient  
 PT infected with P carinii

PS Claim 3; Fig 3; 42pp; English.

XX The amino acid sequence is that of P. carinii cytochrome oxidase

CC subunit II which was translated from DNA from plasmid pA6112. See

CC also AAR21409-R21413 and AAR20056.

XX Sequence 243 AA;

Query Match 27.9%; Score 62; DB 13; Length 243;  
 Best Local Similarity 31.7%; Pred. No. 0.39;  
 Matches 13; Conservative 8; Mismatches 20; Indels 0; Gaps 0;



QY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPEDGIVNEG 42  
 Db 73 lftwtvspalllatafapstfkllymdeviapsitkaig 113

RESULT 4  
 AAW20218  
 ID AAW20218 standard; Protein: 154 AA.  
 AC AAW20218;  
 XX  
 XX 30-JUL-1997 (first entry)

XX H. pylori secreted or periplasmic protein 21687842.aa.  
 DE Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 XX binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 KW Helicobacter pylori.  
 XX  
 OS Location/Qualifiers

XX Key  
 FH Misc-difference 56 /label= "unknown"  
 FT /note= "encoded by RAR"  
 FT Misc-difference 57 /label= "unknown"  
 FT /note= "encoded by GRC"  
 FT Misc-difference 99 /label= "unknown"  
 FT /note= "encoded by TRP"  
 FT Misc-difference 102 /label= "unknown"  
 FT /note= "encoded by TRP"  
 FT Misc-difference 107 /label= "unknown"  
 FT /note= "encoded by TMC"  
 FT Misc-difference 114 /label= "unknown"  
 FT /note= "encoded by ART"  
 FT Misc-difference 120 /label= "unknown"  
 FT /note= "encoded by RAR"  
 FT Misc-difference 126 /label= "unknown"  
 FT /note= "encoded by GRC"

XX WO9640893-A1.

XX 19-DEC-1996.  
 PD 06-JUN-1996; 96WO-US09122.  
 PF

XX 01-APR-1996; 96US-0630405.  
 PR 07-JUN-1995; 95US-0487032.

XX (ASTR ) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaerd BL;

DR WPI: 1997-052306/05.  
 N-PSDB: AAT67749.

XX Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter

XX Claim 72; Page 431; 1481pp; English.

CC This sequence is a H. pylori secreted or periplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. No  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.  
 CC  
 SQ Sequence 154 AA;

Query Match 27.0%; Score 60; DB 18; Length 154;  
 Best Local Similarity 40.0%; Pred. No. 0.45; 10; Indels 0; Gaps 0;  
 Matches 12; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 LEIWTIVPVILVIFGAFSLPVLFKQEP 31  
 Db 1 mkivslvalvfcclgavelpyyqtgef 30

RESULT 5  
 AAW24610  
 ID AAW24610 standard; Protein: 154 AA.

XX AAW24610;  
 AC 11-AUG-1997 (first entry)  
 DT  
 XX  
 DE H. pylori secreted or periplasmic protein 21687842.aa.

XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;  
 KW detection; antisense; inhibition.

XX Helicobacter pylori.

XX Key Location/Qualifiers  
 FH Misc-difference 56 /note= "encoded by RAR"  
 FT Misc-difference 57 /note= "encoded by GRC"  
 FT Misc-difference 99 /note= "encoded by TRP"  
 FT Misc-difference 102 /note= "encoded by TRP"  
 FT Misc-difference 107 /note= "encoded by TMC"  
 FT Misc-difference 114 /note= "encoded by ART"  
 FT Misc-difference 120 /note= "encoded by RAR"  
 FT Misc-difference 126 /note= "encoded by GRC"

XX WO9719098-A1.

XX 29-MAY-1997.

XX 15-NOV-1996; 96WO-US18542.

XX 17-NOV-1995; 95US-0561469.

XX (ASTR ) ASTRA AB.

XX Smith DH;

DR WPI: 1997-298052/27.  
 DR N-PSDB: AAT77428.

XX Helicobacter pylori nucleic acid sequences and related proteins -  
PT used for diagnostics and therapeutics  
XX  
XX  
PS Claim 10; Page 146-147; 1481pp; English.

CC This sequence is a H. pylori secreted or periplasmic protein.  
CC Helicobacter pylori has been strongly linked to chronic gastritis and  
CC duodenal ulcer disease. The nucleic acid sequences of the invention  
CC are used to evaluate compounds, especially activators or inhibitors of  
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
CC sequence. The nucleic acid sequences, and corresponding proteins, are  
CC also useful for generating vaccines for immunizing subjects against H.  
CC pylori or for use in detecting the presence of Helicobacter species in  
CC a sample. Antisense nucleic acid sequences of these sequences are  
CC used to inhibit expression of a gene from Helicobacter species. H.  
CC pylori whole genomic DNA was isolated and nebulized to a median size of  
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
CC complementary to the BstXI-cut PMPX vectors, while the overhang is not  
CC self-complementary. Therefore the linkers will not concatamerise nor  
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts  
CC were ligated to each of the 20 PMPX vectors to construct a series of  
CC shotgun subclone libraries. The purified DNA samples were then  
CC sequenced.  
CC Note: The ORF/protein reference number for this sequence was obtained  
CC from the related specification, WO9640893.

Sequence 154 AA:

Query Match 27.0%; Score 60; DB 18; Length 154;  
Best Local Similarity 40.0%; Pred. No. 0.45; Mismatches 12; Conservative 8; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIFGAFSLPVLFKQOE 31  
DB 1 mkIvslIvavIvfcclIgvavelpgyvtqetf 30

RESULT 6

AAV10989  
ID AAV10989 standard; Protein; 167 AA.

XX AAV10989;

DT 08-JUN-1999 (first entry)

DE H. pylori ORF 06ep11202\_21687842\_c3\_35 secreted protein.

KW Vaccine; probe; diagnostic; ORF; cell envelope protein;

XX secreted protein; cellular protein.

OS Helicobacter pylori.

XX WO9818323-A1.

PN 07-MAY-1998.

PD 28-OCT-1997; 97WO-US19575.

PF 14-JUL-1997; 97US-0891928.

PR 28-OCT-1996; 96US-0739150.

PR 06-DEC-1996; 96US-0739739.

XX (ASTR ) ASTRA AB.

XX Alm RA, Smith D;

XX WPI: 1998-271811/24.

DR N-PSDB; AAX30456.

XX Helicobacter pylori nucleic acids and proteins - used to develop

PT products for the detection, prevention and treatment of H. pylori  
PT infections  
XX  
XX  
PS Claims 27, 31; Page 194; 279pp; English.

CC Recombinant or substantially pure preparations of H. pylori polypeptides  
CC are disclosed, together with the nucleic acids encoding them. In all,  
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
CC secreted proteins or other cellular proteins. Vaccines containing the  
CC nucleic acids or proteins are claimed, as are probes containing at least  
CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
CC for treating or reducing the risk of H. pylori infections, and the  
CC probes can be used diagnostically for detecting the presence of  
CC Helicobacter in a sample. The products are also of use in screening  
CC for compounds having the ability to interfere with the H. pylori life  
CC cycle or to inhibit H. pylori infection.

Sequence 167 AA:

Query Match 27.0%; Score 60; DB 19; Length 167;  
Best Local Similarity 40.0%; Pred. No. 0.49; Mismatches 12; Conservative 8; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIFGAFSLPVLFKQOE 31  
DB 1 mkIvslIvavIvfcclIgvavelpgyvtqetf 30

RESULT 7

AAW20967  
ID AAW20967 standard; Protein; 181 AA.

XX AAW20967;

DT 21-JUL-1997 (first entry)

DE H. pylori secreted or periplasmic protein, hp2el1858orf5.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacterium; life cycle; activator;

KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;

XX diagnosis.

OS Helicobacter pylori.

XX WO9640893-A1.

PN 19-DEC-1996.

PF 06-JUN-1996; 96WO-US09122.

PR 01-APR-1996; 96US-0630405.

PR 07-JUN-1995; 95US-0487032.

XX (ASTR ) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI: 1997-052306/05.

DR N-PSDB; AAT68220.

XX Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

XX infection, and to detect Helicobacter

PS Claim 72; Page 1358; 1481pp; English.

CC This sequence represents a H. pylori secreted or periplasmic protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from

overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified CC and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.

Sequence 181 AA;

Query Match 27.0%; Score 60; DB 18; Length 181;  
Best Local Similarity 40.0%; Pred. No. 0.54;  
Matches 12; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGASLPLVLFKQDEF 31  
DB 8 mklvalvalvfccflgavelpgyvqlqet 37

# RESULT 8

AAV97755 AAV97755 standard; Peptide: 20 AA.

AAV97755;

06-AUG-2001 (first entry)

G. oxydans cytochrome C oxidase (COII) protein sequence fragment.

Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;  
oxidative fermentation; electron transfer; respiratory chain; L-sorbose;  
2-keto-L-gluconic acid; 2KGA production; aldehyde production;  
carboxylic acid production; ketone production.

Gluconobacter oxydans.

EP103603-A2.

30-MAY-2001.

14-NOV-2000; 2000EP-0124785.

17-NOV-1999; 99EP-0122842.

(HOFF) HOFFMANN LA ROCHE & CO AG F.

Asakura A, Hoshino T, Shinjoh M;

WPI; 2001-357953/38.

New cytochrome c oxidase complex having cytochrome c oxidase activity  
from Gluconobacter oxydans DSM 4025, useful in mediating electron  
transfer in respiratory chain or producing 2-keto-L-gluconic acid from  
L-sorbose or D-sorbitol

Example 2; Page 9; 42pp; English.

This sequence represents a fragment of the Gluconobacter oxydans  
cytochrome C oxidase, COII. The invention relates to a cytochrome c  
oxidase COI complex. The COI complex is useful in improving oxidative  
fermentation and is an essential component mediating electron transfer  
in the respiratory chain. The recombinant microorganism and the  
cytochrome c oxidase may be used in the genetic preparation of a  
recombinant COI complex and in the production of 2-keto-L-gluconic acid  
(2KGA) from L-sorbose or D-sorbitol in a culture medium. The COI is also  
useful as a terminal oxidase, oxidising cytochrome c, an electron  
acceptor from an enzyme belonging to dehydrogenase for the production of  
aldehydes, carboxylic acids and ketones from alcohols and aldehydes,  
especially the production of 2KGA from L-sorbose or D-sorbitol.

Sequence 20 AA;

Query Match 26.6%; Score 59; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PLEWTIVPV 11  
DB 10 plewtivpv 20

# RESULT 9

AAG03705 AAG03705 standard; Protein: 55 AA.

AAG03705;

06-OCT-2000 (first entry)

Human secreted protein, SEQ ID NO: 7786.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
gene therapy; chromosome mapping.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

N-PSDB; AAG03711.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
diagnostic, forensic, gene therapy and chromosome mapping procedures -  
Claim 13; SEQ ID 7786; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number  
of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
were prepared from total human RNAs or polyA+ RNAs derived from 30  
different tissues. EST sequences usually correspond mainly to the 3'  
untranslated region (UTR) of the mRNA because they are often obtained  
from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
those cases where longer cDNA sequences have been obtained, the full 5'  
UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
ends and can therefore be used to obtain full length cDNAs and genomic  
DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
chromosome mapping procedures. They are used to obtain upstream  
regulatory sequences and to design expression and secretion vectors.

Sequence 55 AA;

Query Match 26.6%; Score 59; DB 21; Length 55;  
Best Local Similarity 42.4%; Pred. No. 0.19;  
Matches 14; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

OY 7 TTPVVIIVIFGASLPLVLFKQDEPGDIVIN 39  
DB 20 tsvpvlfcvcrkpvlysqllstqstivln 52

## RESULT 10

AAG90866  
ID AAG90866 standard; Protein; 197 AA.

XX  
AC AAG90866;

XX  
DT 26-SEP-2001 (first entry)

XX  
DE C glutamicum protein fragment SEQ ID NO: 4620.

XX  
KW Corynebacterium: amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis.

XX  
OS Corynebacterium glutamicum.

XX  
PN EPI108790-A2.

XX  
PD 20-JUN-2001.

XX  
PF 18-DEC-2000; 2000EP-0127688.

XX  
PR 16-DEC-1999; 99JP-0377484.

XX  
PR 07-APR-2000; 2000JP-0159162.

XX  
PR 03-AUG-2000; 2000JP-0280988.

XX  
PA (KYOW ) KYOWA HAKKO KOCYO KK.

XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX  
DR N-PSDB; AAH66085.

XX  
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying

XX  
PT mutation point of a gene, measuring expression of a gene, analysing

XX  
PT expression profile or pattern of a gene and identifying homologous gene

XX  
PS Claim 17; SEQ ID NO: 4620; 246pp + Sequence Listing; English.

XX  
CC The present invention provides a number of nucleotide and protein

XX  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These

XX  
CC are useful for identifying the mutation point of a gene derived from a

XX  
CC mutant of corynebacterium bacterium, measuring expression amount and

XX  
CC analysing the expression profile or expression pattern of a gene derived

XX  
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived

XX  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing

XX  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

XX  
CC particularly L-lysine. The present sequence is a protein described

XX  
CC in the exemplification of the invention.

XX  
CC Note: The sequence data for this patent did not form part of the printed

XX  
CC specification, but was obtained in electronic format directly from the

XX  
CC European Patent Office.

XX  
SQ Sequence 197 AA.

Query Match 26.6%; Score 59; DB 22; Length 197;

Best Local Similarity 33.3%; Pred. No. 0.84;

Matches 19; Conservative 9; Mismatches 11; Indels 18; Gaps 3;

QY 4 IWTIVPVVILVIFGAFS---LPV-----LFKQOE-----FPGDIVINVEG 42

DB 120 IYIGTAPAILLVFAGAGFGLVPGFTLMIYVAFIRGDKLLKYGKYPVWILVGVIG 176

## RESULT 11

AAE05103

XX ID AAE05103 standard; Protein; 593 AA.

XX AC AAE05103;

XX XX

DT 18-SEP-2001 (first entry)

XX  
DE Drosophila melanogaster dmsNF2.

XX  
KW Fruit fly; invertebrate symporter cell surface receptor; dmsNF2;

XX  
KW sodium/neurotransmitter family; SNF; biopesticide; therapy.

XX  
OS Drosophila melanogaster.

XX  
PN WO200149848-A2.

XX  
PD 12-JUL-2001.

XX  
PF 28-DEC-2000; 2000WO-US35551.

XX  
PR 30-DEC-1999; 99US-0173929.

XX  
PR 15-MAR-2000; 2000US-0189399.

XX  
PR 23-MAR-2000; 2000US-0191686.

XX  
PR 23-MAR-2000; 2000US-0191687.

XX  
PR 23-MAR-2000; 2000US-0191688.

XX  
PR 23-MAR-2000; 2000US-0191695.

XX  
PA (GENO-) GENOPTERA LLC.

XX  
PI Kellerman KA, Keegan KP, Ebens AJ, Torpey J;

XX  
PI WPI: 2001-441879/47.

XX  
DR N-PSDB; AAD09683.

XX  
PT Novel invertebrate symporter cell surface receptor proteins and nucleic

XX  
PT acid encoding the protein useful as pesticide or drug target and to

XX  
PT identify compounds that have utility as therapeutics or pesticides

XX  
PS Claim 18; Page 65-66; 71pp; English.

XX  
CC The invention relates to invertebrate symporter cell surface receptors

XX  
CC of the sodium/neurotransmitter family (SNF) and nucleic acid molecules

XX  
CC encoding such receptors. The SNF protein is useful for detecting a

XX  
CC candidate compound especially a putative pesticide or pharmaceutical

XX  
CC agent that interacts with an invertebrate symporter cell surface

XX  
CC receptor protein or its fragment. Insect or worm genetically modified

XX  
CC symporter cell surface receptor protein activity, by detecting the

XX  
CC phenotype caused by the expression or mis-expression of the protein in

XX  
CC the animal. Nucleic acids encoding the invertebrate receptor protein or

XX  
CC their fragments are useful as biopesticides. SNF nucleic acids are

XX  
CC useful for generating mutant phenotypes in an animal model or living

XX  
CC cells that are used to study the regulation of genes encoding the

XX  
CC modified organisms or cells are useful in screening assays to identify

XX  
CC pesticides or therapeutics and thus are useful in the identification of

XX  
CC new drug targets, therapeutic agents, diagnostics and prognostics

XX  
CC useful in treatment of disorders associated with ion channels. The

XX  
CC nucleic acid molecules are also useful as hybridisation probes. The

XX  
CC The present sequence is Drosophila melanogaster (dm) SNF homologue

XX  
CC which is referred as dmsNF2.

SQ Sequence 593 AA;

Query Match 25.5%; Score 56.5; DB 22; Length 593;

Best Local Similarity 34.9%; Pred. No. 7.2;

Matches 15; Conservative 9; Mismatches 12; Indels 7; Gaps 2;

QY 4 IWTIVPVVILVIFGAFSL---PVLFKQOEFPFGDIVINVEG 42

DB 472 IOWSVLPLVLMVLVYLSLLCMPLSYNGELP---LVYRVG 511

## RESULT 12

AAG92151

XX ID AAG92151 standard; Protein; 359 AA.

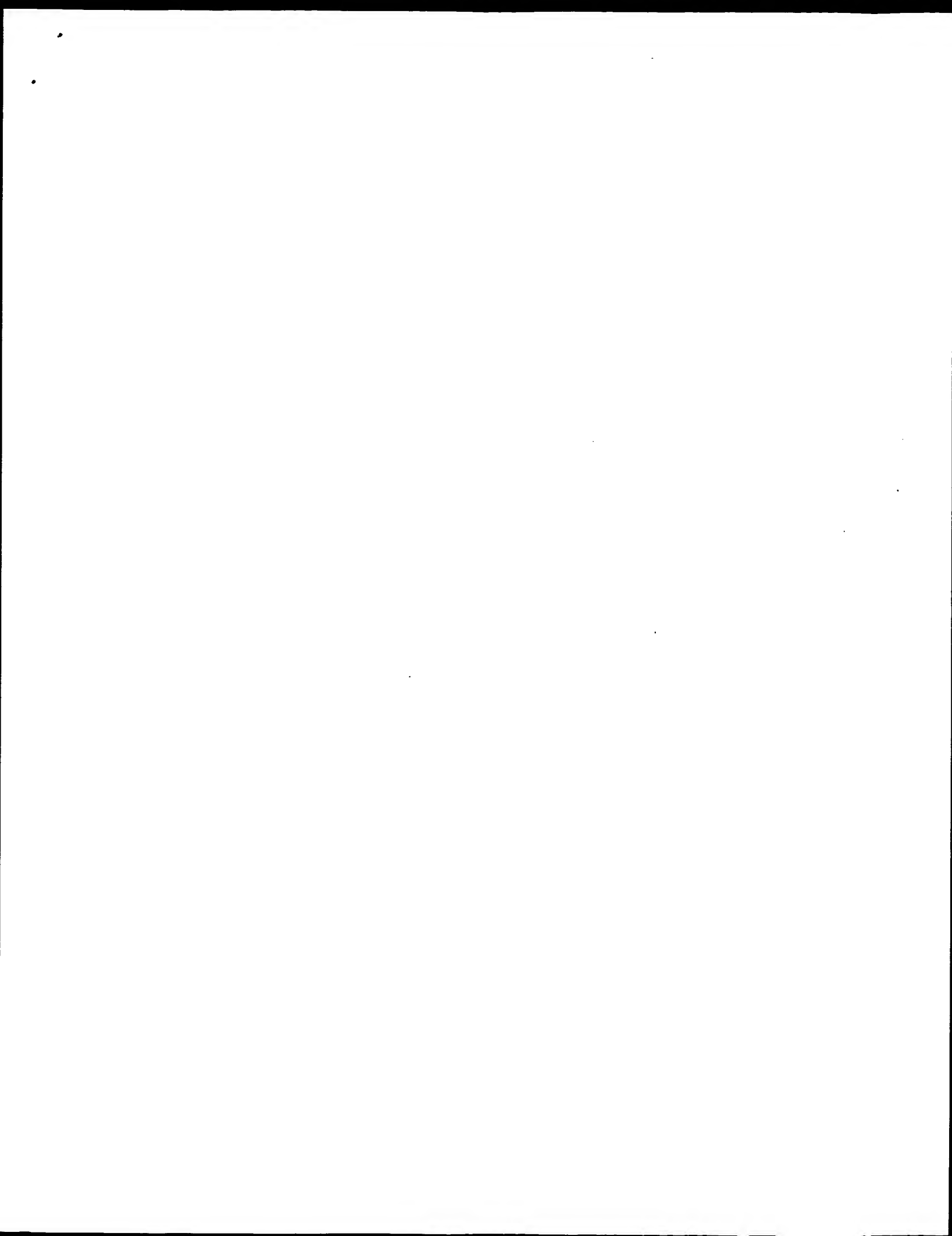
XX XX

AC	AAG92151;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum protein fragment SEQ ID NO: 5905.
XX	
KW	Corynebacterium; amino acid synthesis; vitamin; saccharide; organic acid synthesis.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EPI108790-A2.
XX	
PD	20-JUN-2001.
XX	
PF	18-DEC-2000; 2000EP-0127688.
XX	
PR	16-DEC-1999; 99JP-0377484. 07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
XX	
PA	(KXOW ) KYOWA HAKKO KOGYO KK.
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
DR	WPI; 2001-376931/40.
XX	
N-PSDB; AAB67370.	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene
PS	
XX	
Claim 17; SEQ ID NO: 5905; 246bp + Sequence Listing; English.	
XX	
CC	The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.
XX	
CC	
XX	
Seq	Sequence 359 AA:
Query Match	25.2%; Score 56; DB 22; Length 359;
Best Local Similarity	45.5%; Pred. NO. 4.77.
Matches 10; Conservative	8; Mismatches 4; Indels 0; Gaps 0;
OY	1 PLEIWTIVPVVILVFIGAFLS 22           ::: :   : Db 108 plelvtivpvilvmlyfftv 129
RESULT 13	
AAB79573	
ID	AAB79573 standard; Protein; 359 AA.
XX	
AC	AAB79573;
XX	
DT	30-APR-2001 (first entry)
XX	
DE	Corynebacterium glutamicum SMP protein sequence SEQ ID NO:662.
XX	
KW	Corynebacterium glutamicum; carbon metabolism and energy production;

KW	SMP protein; sugar metabolism and oxidative phosphorylation protein;
KV	flame chemical production; organic acid; proteinoogenic amino acid;
KW	noproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KM	nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW	carbohydrate; aromatic compound; vitamin; cofactor; polyeptide; enzyme;
KX	diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX	Corynebacterium glutamicum.
OS	MO200100844-A2.
PN	04-JAN-2001.
PD	23-JUN-2000; 2000WO-IB00943.
PE	
XX	
XX	25-JUN-1999; 99US-0141031.
PR	08-JUL-1999; 99DE-1031412.
PR	08-JUL-1999; 99DE-1031413.
PR	08-JUL-1999; 99DE-1031419.
PR	08-JUL-1999; 99DE-1031420.
PR	08-JUL-1999; 99DE-1031424.
PR	08-JUL-1999; 99DE-1031428.
PR	08-JUL-1999; 99DE-1031431.
PR	08-JUL-1999; 99DE-1031433.
PR	08-JUL-1999; 99DE-1031434.
PR	08-JUL-1999; 99DE-1031510.
PR	08-JUL-1999; 99DE-1031562.
PR	08-JUL-1999; 99DE-1031634.
PR	09-JUL-1999; 99DE-1032180.
PR	09-JUL-1999; 99DE-1032227.
PR	09-JUL-1999; 99DE-1032230.
PR	09-JUL-1999; 99US-0143208.
PR	14-JUL-1999; 99DE-1032924.
PR	14-JUL-1999; 99DE-1032973.
PR	27-AUG-1999; 99DE-1033005.
PR	31-AUG-1999; 99DE-1040765.
PR	03-SEP-1999; 99US-0151572.
PR	03-SEP-1999; 99DE-1042076.
PR	03-SEP-1999; 99DE-1042079.
PR	03-SEP-1999; 99DE-1042086.
PR	03-SEP-1999; 99DE-1042087.
PR	03-SEP-1999; 99DE-1042088.
PR	03-SEP-1999; 99DE-1042095.
PR	03-SEP-1999; 99DE-1042123.
PR	03-SEP-1999; 99DE-1042125.
PA	(BAD1 ) BASF AG.
XX	
PI	Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
DR	WPI: 2001-061975/07.
XX	
DR	N-P-SDB; AAF71690.
XX	
PT	New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT	metabolism and oxidative phosphorylation protein for production or
PT	modulation of production of fine chemicals e.g. amino acids,
XX	carbohydrates or enzymes -
PS	Claim 20; Page 1081-1082; 1246pp; English.
CC	AAE71360 to AAE71750 encode the Corynebacterium glutamicum sugar
CC	metabolism and oxidative phosphorylation (SMP) proteins given in
CC	AAE79243 to AAB 79633 which are involved in carbon metabolism and
CC	energy production. The C. glutamicum SMP gene can be used in vectors
CC	(II) for expression in host cells and production or modulation of
CC	production of fine chemicals, such as, an organic acid, a proteinoenic
CC	or nonproteinoenic amino acid (preferred), a purine or pyrimidine base,
CC	a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC	acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC	cofactor, a polyeptide, or an enzyme. The presence of (I) or SMP proteins
CC	(III) encoded by them are used for diagnosing the presence or activity of
CC	Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC	containing them are used to map genomes of organisms related to











Db 61 METWTLPALILVILALPSRLILWTFDEVNDPSLTKSIG 101

## RESULT 2

US-08-415-751-7

Sequence 7, Application US/08415751

Patent No. 5643772

GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEECH, JAMES

APPLICANT: NELSON, RICHARD, C.

TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-

TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA

TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID

TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND

TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESS: PHILLIPS, MOORE, LEMPIO &amp; FINLEY

STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto

STATE: California

COUNTRY: United States of America

ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage

COMPUTER: PC

OPERATING SYSTEM: DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/415,751

FILING DATE: 03-Apr-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/071,880

FILING DATE: June 1, 1993

APPLICATION NUMBER: 07/891,301

FILING DATE: May 29, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Hana Dolezalova

REGISTRATION NUMBER: 30,518

REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-1677

TELEFAX: (415) 324-1678

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 361 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Cryptosporidium parvum

FEATURE:

NAME/KEY: Positions coded by nonsense codons are

NAME/KEY: Identified as Xaa.

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

US-08-415-751-7

Query Match 23.9%; Score 53; DB 2; Length 322;  
Best Local Similarity 48.0%; Pred. No. 6.2;  
Matches 12; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 LEIVTVIVVILVIGAFSLPVLF 26  
DB 145 MEIRITFSPVILFIAIPSFALLY 169

RESULT 5  
US-07-759-568-1

; Sequence 1, Application US/07759568  
; Patent No. 5374506  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Philip M.  
; TITLE OF INVENTION: Cloning of cDNA Encoding a Functional  
; TITLE OF INVENTION: Human Interleukin-8 Receptor  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cushman, Darby & Cushman  
; STREET: 1615 L Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/759,568  
; FILING DATE: 19910913  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Scott, Watson T.  
; REGISTRATION NUMBER: 26581  
; REFERENCE/DOCKET NUMBER: WTS/5683/91535/WBH  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3000  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 cush  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 355 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-759-568-1

Query Match 22.5%; Score 50; DB 1; Length 355;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIGAFSLPVLF 27  
DB 216 IVPILIMFCYGFLLRLTLFK 235

RESULT 6  
US-08-450-393A-8

; Sequence 8, Application US/08450393A  
; Patent No. 5707815  
; GENERAL INFORMATION:  
; APPLICANT: Charo, Israel  
; APPLICANT: Coughlin, Shaun  
; TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOTACTANT  
; TITLE OF INVENTION: PROTEIN RECEPTORS  
; NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
STREET: 5 Palo Alto Square  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94306-2155

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450,393A

FILING DATE: May 25, 1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Cserf, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: UCAL-237/0205

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-843-5165

TELEFAX: 415-887-0663

TELEX: 380816COOLEYPA

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-450-393A-8

Query Match 22.5%; Score 50; DB 1; Length 355;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIGAFSLPVLF 27  
DB 216 IVPILIMFCYGFLLRLTLFK 235

RESULT 7

US-08-390-000A-5

; Sequence 5, Application US/08390000A

; Patent No. 5985583

; GENERAL INFORMATION:

; APPLICANT: Sealton, Stuart C.

; TITLE OF INVENTION: Cloning and Expression of

; TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor

; NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Penile & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/390,000A

FILING DATE: 17-FEB-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Miscock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 6923-052

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 355 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-390-000A-5

Query Match 22.5%; Score 50; DB 2; Length 355;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIFGAFSLPVLFK 27  
|||:::| 1:1 |||  
Db 216 IVPILIMFCYGTLLRTLFK 235

RESULT 8  
US-08-446-669-8  
Sequence 8, Application US/08446669  
Patent No. 6132987  
GENERAL INFORMATION:  
APPLICANT: Chemo, Israel  
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOATTRACTANT  
NUMBER OF INVENTIONS: PROTEIN RECEPTORS  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
STREET: 5 Palo Alto Square  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94306-2155  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,669  
FILING DATE: May 25, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Neeley, Richard  
REGISTRATION NUMBER: 30,092  
REFERENCE/DOCKET NUMBER: UCAL-237/01US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-843-5000  
TELEFAX: 415-857-0663  
TELEX: 380816COOLEYPA  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 355 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-446-669-8

Query Match 22.5%; Score 50; DB 4; Length 355;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIFGAFSLPVLFK 27  
|||:::| 1:1 |||  
Db 216 IVPILIMFCYGTLLRTLFK 235

RESULT 9  
PCT-US95-00476-8  
Sequence 8, Application PC/TUS9500476  
GENERAL INFORMATION:  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOATTRACTANT  
NUMBER OF INVENTIONS: PROTEIN RECEPTORS  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Berliner & Carson  
STREET: 201 N. Figueroa Street, 5th Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90012-2628  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/00476  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Berliner, Robert  
REGISTRATION NUMBER: 20,121  
REFERENCE/DOCKET NUMBER: 5555-291  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 310-977-1001  
TELEFAX: 310-977-1003  
TELEX:  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 355 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
PCT-US95-00476-8

Query Match 22.5%; Score 50; DB 5; Length 355;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIFGAFSLPVLFK 27  
|||:::| 1:1 |||  
Db 216 IVPILIMFCYGTLLRTLFK 235

RESULT 10  
US-08-202-056-7  
Sequence 7, Application US/08202056  
Patent No. 5440021  
GENERAL INFORMATION:  
APPLICANT: Chuntharapal, Anan  
APPLICANT: Hebert, Caroline  
APPLICANT: Kim, Kyung Jin  
APPLICANT: Lee, James  
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible





Thu Dec 27 08:21:48 2001

us-09-712-768-4.ra1





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:02:25 ; Search time 24.27 Seconds  
(Without alignments)  
138,100 Million cell updates/sec

Title: US-09-712-768-4  
Perfect score: 222  
Sequence: 1 PLEIWTIVPVILVIFGAF.....LFKQDFPGDIVINVEGRS 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159	71.6	302	2 J01013	cytochrome-c oxida
2	139.5	62.8	297	1 OBPC2N	cytochrome-c oxida
3	96	43.2	313	2 E71698	cytochrome-c oxida
4	90	40.5	253	2 T12405	cytochrome-c oxida
5	89	40.1	229	2 S50328	cytochrome-c oxida
6	87	39.2	227	2 T11183	cytochrome-c oxida
7	86	38.7	227	2 S10190	cytochrome-c oxida
8	86	38.7	227	2 T11327	cytochrome-c oxida
9	86	38.7	227	2 T11025	cytochrome-c oxida
10	85	38.3	227	2 T11078	cytochrome-c oxida
11	85	38.3	229	2 T11170	cytochrome-c oxida
12	85	38.3	229	2 T12412	cytochrome-c oxida
13	84	37.8	230	2 S36009	cytochrome-c oxida
14	83	37.4	229	2 T11796	cytochrome-c oxida
15	83	37.4	230	2 T11537	cytochrome-c oxida
16	83	37.4	230	2 T11290	cytochrome-c oxida
17	82	36.9	227	2 T11196	cytochrome-c oxida
18	82	36.9	230	2 S68131	cytochrome-c oxida
19	82	36.9	230	2 T11767	cytochrome-c oxida
20	82	36.9	230	2 T09860	cytochrome-c oxida
21	82	36.9	230	2 T09950	cytochrome-c oxida
22	82	36.9	230	2 T11303	cytochrome-c oxida
23	81	36.5	225	2 T11155	cytochrome-c oxida
24	81	36.5	225	2 S70599	cytochrome-c oxida
25	81	36.5	230	2 S45491	cytochrome-c oxida
26	81	36.5	232	2 T11823	cytochrome-c oxida
27	81	36.5	232	2 T11893	cytochrome-c oxida
28	79	35.6	228	2 C38941	cytochrome-c oxida
29	79	35.6	228	2 T11130	cytochrome-c oxida

30	79	35.6	229	2 A45170	cytochrome-c oxida
31	79	35.6	231	2 184422	cytochrome-c oxida
32	79	35.6	234	2 S35465	cytochrome-c oxida
33	79	35.6	237	2 S26949	cytochrome-c oxida
34	78	35.1	224	2 F45170	cytochrome-c oxida
35	78	35.1	228	1 OBF2	cytochrome-c oxida
36	78	35.1	228	1 OBF2Y	cytochrome-c oxida
37	78	35.1	228	1 T12000	cytochrome-c oxida
38	78	35.1	228	1 T36906	cytochrome-c oxida
39	77	34.7	226	2 T11352	cytochrome-c oxida
40	77	34.7	227	2 T62728	cytochrome-c oxida
41	77	34.7	230	2 T11457	cytochrome-c oxida
42	77	34.7	258	2 T11923	cytochrome-c oxida
43	77	34.7	873	1 S53828	cytochrome-c oxida
44	76	34.2	228	2 B38941	cytochrome-c oxida
45	76	34.2	228	2 T09802	cytochrome-c oxida

## ALIGNMENTS

RESULT 1 J01013 cytochrome-c oxidase (EC 1.9.3.1) chain II precursor - Rhodobacter sphaeroides

C:Species: Rhodobacter sphaeroides

C:Date: 23-Nov-1991 #sequence\_revision 23-Nov-1991 #text\_change 23-Jul-1999

C:Accession: J01013

R:Gao, J.; Shapleigh, J.; Gennis, R.; Revzin, A.; Ferguson-Miller, S.

Gene 101, 133-137, 1991

A:Title: The gene encoding cytochrome c oxidase subunit II from Rhodobacter sphaeroid

A:Reference number: J01013; MUID:91285423

A:Accession: J01013

A:Molecule type: DNA

A:Residues: 1-302 <CAO>

A:Cross-references: GB:M57680; NID:g151893; PIDN:AAA26100.1; PID:g151894

C:Genetics:

A:Gene: coxII

C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology

C:Keywords: copper; electron transfer; heme; membrane-associated complex; oxidoreduct

F:1-24/Domain: (or 1-25) signal sequence #status predicted <SIG>

F:25-302/Product: (or 26-302) cytochrome-c oxidase polypeptide II #status predicted <

F:43-269/Domain: cytochrome-c oxidase chain II homology <CO2>

F:217-251,255-269/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted

F:251,253,255,259/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted

F:253/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 71.6%; Score 159; DB 2; Length 302;

Best Local Similarity 73.8%; Pred. No. 1,4e-12;

Matches 31; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PLEIWTIVPVILVIFGAFSLVLFKQDFPGDIVINVEG 42

Db 99 PLEIWTIVPVILVIFGAFSLVLFKQDFPGDIVINVEG 140

RESULT 2 OBPC2N cytochrome-c oxidase (EC 1.9.3.1) chain II precursor - Paracoccus denitrificans

C:Species: Paracoccus denitrificans

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999

C:Accession: S00106; S20231; S03803

R:Steinmeyer, P.; Steffens, G.C.M.; Panskus, G.; Buse, G.; Ludwig, B.

A:Title: Subunit II of cytochrome c oxidase from Paracoccus denitrificans. DNA sequen

A:Reference number: S00106; MUID:88004464

A:Accession: S00106

A:Molecule type: DNA

A:Residues: 1-297 <STE>

A:Cross-references: EMBL:X05934; NID:g45483; PIDN:CAA29372.1; PID:g45484

A:Accession: S20231

A:Molecule type: protein

A:Residues: 29; B:261-280 <STE2>

R:Raitio, M.; Jalli, T.; Saraste, M.  
EMBO J. 6, 2825-2833, 1987  
A:Title: Isolation and analysis of the genes for cytochrome c oxidase in *Paracoccus denitrificans*  
A:Reference number: S03803  
A:Accession: S03803  
A:Molecule type: DNA  
A:Residues: 1-158, 'GV', 161-297 <RA1>  
A:Cross-references: EMBL:X05828; NID:g45468; PIDN:CAA29268.1; PID:g45469  
C:Genetics:  
A:Gene: COII  
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; membrane-associated complex; oxidoreductase; pyro  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-280/Product: cytochrome-c oxidase chain II #status experimental <MAT>  
F:29-54/Domain: periplasmic #status predicted <PER1>  
F:48-62/Domain: cytochrome-c oxidase chain II homology <CO2>  
F:55-87/Domain: transmembrane #status experimental <TM1>  
F:102-133/Domain: transmembrane #status experimental <TM2>  
F:134-280/Domain: carboxyl-terminal propetide #status predicted <CTP>  
F:281-297/Domain: carboxyl-terminal propetide #status predicted <CTP>  
F:29/Modified site: pyrrolidone carboxylic acid (Glu) (in mature form) #status experim  
F:209,244,248,255/Binding site: copper 1 (His, Cys, Cys, Met) #status experimental  
F:244,246,248,252/Binding site: copper 2 (Cys, Glu, Cys, His) #status experimental  
F:246/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 62.8% Score 139.5; DB 1; Length 297;  
Best Local Similarity 60.5% Pred. No. 3,8e-10;  
Matches 26; Conservative 9; Mismatches 7; Indels 1; Gaps 1;  
OY 1 LEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 42  
DB 104 PREIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 146

RESULT 3  
E71698  
Cytochrome-c oxidase (EC 1.9.3.1) chain II RP406 - *Rickettsia prowazekii*  
C:Species: *Rickettsia prowazekii*  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: E71698  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: E71698  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-313 <AND>  
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14863.1; PID:g386096  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: COXB; RP406  
C:Superfamily: cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; membrane-associated complex; oxidoreductase; resp  
F:76-286/Domain: cytochrome-c oxidase chain II homology <CO2>  
F:233,268,272,279/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
F:268,270,272,276/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F:270/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 43.2% Score 96; DB 2; Length 313;  
Best Local Similarity 41.5% Pred. No. 0.00012;  
Matches 17; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
OY 2 LEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 42  
DB 133 IEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 173

RESULT 4  
T12405  
Cytochrome-c oxidase (EC 1.9.3.1) chain II - *Sarcophyton glaucum* mitochondrion

C:Species: mitochondrion *Sarcophyton glaucum*  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 21-Jan-2000  
C:Accession: T12405  
R:Beaton, M.J.; Roger, A.J.; Cavalier-Smith, T.  
J. Mol. Evol. 47, 697-708, 1998  
A:Title: Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*: Conser  
A:Reference number: Z17505; MUID:99065763  
A:Accession: T12405  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-253 <BEA>  
A:Cross-references: EMBL:AF064823; NID:g4091912; PID:g4091916; PIDN:AAC99650.1  
C:Genetics:  
A:Genome: mitochondrion  
A:Note: COII  
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inn  
F:25-237/Domain: cytochrome-c oxidase chain II homology <CO2>  
F:184,219,223,230/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
F:219,221,223,227/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F:221/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 40.5% Score 90; DB 2; Length 253;  
Best Local Similarity 43.9% Pred. No. 0.00053;  
Matches 18; Conservative 7; Mismatches 16; Indels 0; Gaps 0;  
OY 2 LEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 42  
DB 77 IEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 117

RESULT 5  
S50328  
Cytochrome-c oxidase (EC 1.9.3.1) chain II - *Katharina tunnicata* mitochondrion  
C:Species: mitochondrion *Katharina tunnicata*  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 07-Dec-1999  
C:Accession: S50328  
R:Boore, J.L.; Brown, W.M.  
Genetics 138, 423-443, 1994  
A:Title: Complete DNA sequence of the mitochondrial genome of the black chiton, *Katha*  
A:Reference number: S50327; MUID:95129806  
A:Accession: S50328  
A:Molecule type: DNA  
A:Residues: 1-229 <BOO>  
A:Cross-references: EMBL:U09810; NID:g557273; PIDN:AAC48365.1; PID:g557277  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC4  
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondri  
F:9-214/Domain: cytochrome-c oxidase chain II homology <CO2>  
F:161,196,200,207/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
F:196,198,200,204/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F:198/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 40.1% Score 89; DB 2; Length 229;  
Best Local Similarity 43.9% Pred. No. 0.00065;  
Matches 18; Conservative 8; Mismatches 15; Indels 0; Gaps 0;  
OY 2 LEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 42  
DB 61 IEIWTIVPVILVFIFGAFSLPVLFKQEPPEGDIVINEG 101

RESULT 6  
T11183  
Cytochrome-c oxidase (EC 1.9.3.1) chain II - *Smithornis sharpei* mitochondrion  
C:Species: mitochondrion *Smithornis sharpei*  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jan-2000  
C:Accession: T11183



```

A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-227 <MIN>
A:Cross-references: EMBL:AF090338; NID:g4894462; PID:g4894466; PIDN:AAD32495.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match          38.3%; Score 85; DB 2; Length 227;
Best Local Similarity 31.7%; Pred. No. 0.002;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFIGAFSLPVLFGQEPEDIVINVEG 42
DB 60 VELIWTILPAIVLILALPSLQILYIMDEIDEDLTAKIG 100

RESULT 11
T11170
Cytochrome-c oxidase (EC 1.9.3.1) chain II - greater rhea mitochondrion
C:Species: mitochondrion Rhea americana (greater rhea, common rhea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T11170; T11419
R:Mindell, D.P.; Sorenson, M.D.; Dincheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: Z17242
A:Accession: T11170
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-229 <MIN>
A:Cross-references: EMBL:AF090339; NID:g4894475; PID:g4894478; PIDN:AAD32505.1
R:Harlid, A.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z17270
A:Accession: T11419
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-229 <HAR>
A:Cross-references: EMBL:Y16884; PIDN:CAA76504.1
C:Genetics:
A:Gene: COII
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial
ein
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match          38.3%; Score 85; DB 2; Length 229;
Best Local Similarity 31.7%; Pred. No. 0.0021;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFIGAFSLPVLFGQEPEDIVINVEG 42
DB 60 VELIWTILPAIVLILALPSLQILYIMDEIDEDLTAKIG 100

RESULT 12
T12412
Cytochrome-c oxidase (EC 1.9.3.1) chain II - ostrich mitochondrion
C:Species: mitochondrion Struthio camelus (ostrich)

```

```

C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C:Accession: T12412; T11522
R:Mindell, D.P.; Sorenson, M.S.; Huddleston, C.J.; Miranda Jr., H.C.; Knight, A.; Saw
submitted to the EMBL Data Library, June 1998
A:Description: Phylogenetic relationships among and within select avian orders based
A:Reference number: Z17507
A:Accession: T12412
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-229 <MIN>
A:Cross-references: EMBL:AF069429; NID:g3283618; PID:g3283625; PIDN:AAD09386.1
R:Harlid, A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z17278
A:Accession: T11522
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-229 <HAR>
A:Cross-references: EMBL:Y12025; PIDN:CAA72747.1
C:Genetics:
A:Gene: COII
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inn
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match          38.3%; Score 85; DB 2; Length 229;
Best Local Similarity 31.7%; Pred. No. 0.0021;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFIGAFSLPVLFGQEPEDIVINVEG 42
DB 60 VELIWTILPAIVLILALPSLQILYIMDEIDEDLTAKIG 100

RESULT 13
S36009
Cytochrome-c oxidase (EC 1.9.3.1) chain II - common carp mitochondrion
C:Species: mitochondrion Cyprinus carpio (common carp)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 23-Mar-2001
C:Accession: S36009; D44650
R:Chang, Y.S.; Huang, F.L.
submitted to the EMBL Data Library, July 1991
A:Description: The cDNA and primary structure of pregrowth hormones of three species
A:Reference number: S21910
A:Accession: S36009
A:Molecule type: DNA
A:Residues: 1-230 <CHAI>
A:Cross-references: EMBL:X61010; NID:g436882; PIDN:CAA43340.1; PID:g12842
R:Chang, Y.S.; Huang, F.L.; Lo, T.B.
J. Mol. Evol. 38, 138-155, 1994
A:Title: The complete nucleotide sequence and gene organization of carp (Cyprinus car
A:Reference number: A44650; MUID:94222691
A:Accession: D44650
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-219, 'F', 221-230 <CHA2>
A:Cross-references: EMBL:X61010; NID:g436882
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial
ein
F:9-214/Domain: cytochrome-c oxidase chain II homology <CO2>
F:161,196,200,207/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:196,198,200,204/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:198/Binding site: magnesium (Glu) (shared with chain I) #status predicted

```

	Query Match	37.8%	Score 84;	DB 2;	Length 230;
	Best Local Similarity	43.9%	Pred. No. 0.0028;		
Matches	18;	Conservative	6;	Mismatches	17; Indels 0; Gaps 0;
Oy	2 LEIWMIVPVVITLPGATSLFKQGEFFPGDGVIVEG	42			
	:     :	:	:	:	:
Db	61 IEIWMIPAVVILPALPSKRIYLMDEINDPHLTIRAMG	101			



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:08:07 : Search time 15.26 Seconds  
(Without alignments)  
105.718 Million cell updates/sec

Title: US-09-712-768-4

Perfect score: 222  
Sequence: 1 PLEIWTIVPVILVFIGNA.....LFKQEPPEGDIVINVEGRS 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	159	71.6	302	1	COX2_RHOSH
2	139.5	62.8	298	1	COX2_PARDE
3	96	43.2	313	1	COX2_RICPR
4	86	38.7	199	1	COX2_APTAU
5	86	38.7	227	1	COX2_CHICK
6	86	38.7	227	1	COX2_COTJA
7	86	38.7	228	1	COX2_ANAPL
8	86	38.7	228	1	COX2_CAIMO
9	85	38.3	198	1	COX2_NOTPE
10	85	38.3	198	1	COX2_TINMA
11	85	38.3	199	1	COX2_CASBE
12	85	38.3	199	1	COX2_DRONO
13	85	38.3	199	1	COX2_RHEAM
14	85	38.3	229	1	COX2_SYRCA
15	84	37.8	230	1	COX2_CARAU
16	84	37.8	230	1	COX2_CYPCA
17	83	37.4	230	1	COX2_SQUAC
18	82	36.9	230	1	COX2_ONCMY
19	82	36.9	230	1	COX2_SALSA
20	82	36.9	230	1	COX2_SCYCA
21	81	36.5	225	1	COX2_RHISA
22	81	36.5	229	1	COX2_ASTPE
23	81	36.5	229	1	COX2_MXGL
24	81	36.5	230	1	COX2_GADMO
25	81	36.5	248	1	COX2_MESE
26	80	36.0	228	1	COX2_ABDAB
27	79	35.6	216	1	COX2_CAGO
28	79	35.6	228	1	COX2_ZOAN
29	79	35.6	229	1	COX2_DROAI
30	79	35.6	229	1	COX2_DROAM
31	79	35.6	229	1	COX2_DROBF
32	79	35.6	229	1	COX2_DROBL
33	79	35.6	229	1	COX2_DRONR

34	79	35.6	229	1	COX2_DROPS	P29864 drosophila
35	79	35.6	229	1	COX2_DROSU	P29865 drosophila
36	79	35.6	230	1	COX2_CROLA	P34189 crossostoma
37	79	35.6	231	1	COX2_LAGIA	P98036 lagothrix 1
38	79	35.6	237	1	COX2_TIRRU	Q01536 trichophyto
39	78	35.1	224	1	COX2_EXERO	P29873 exoristes r
40	78	35.1	228	1	COX2_ANOUP	P33505 anophelies q
41	78	35.1	228	1	COX2_DROME	P00408 drosophila
42	78	35.1	228	1	COX2_DROST	P50253 drosophila
43	78	35.1	229	1	COX2_DROYA	P00409 drosophila
44	78	35.1	228	1	COX2_STMWI	P98021 stimulum vi
45	78	35.1	231	1	COX2_ALOPA	P98024 alouatta pa

## ALIGNMENTS

RESULT	ID	COX2_RHOSH	STANDARD:	PRT:	302 AA.
AC	003736:				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	CYTOCHROME C OXIDASE POLYPEPTIDE II PRECURSOR (EC 1.9.3.1)				
DE	(CYTOCHROME A3 SUBUNIT 2) (OXIDASE AA(3) SUBUNIT 2).				
GN	CTAC OR CTAB OR COXII.				
OS	Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;				
OC	Rhodobacter.				
OX	NCBI_TaxID=1063;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE-91285423; PubMed=1648008;				
KA	Cao J., Shapleigh J., Gennis R., Reyzin A., Ferguson-Miller S.;				
RT	"The gene encoding cytochrome c oxidase subunit II from Rhodobacter				
RT	sphaeroides; comparison of the deduced amino acid sequence with				
RT	sequences of corresponding peptides from other species."				
RL	Gene 101:133-137(1991).				
CC	-1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME				
CC	COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA				
CC	HEME A AND CU(II) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND				
CC	CU(II).				
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +				
CC	4 FERRICYTOCHROME C.				
CC	-1- COFACTOR: COPPER A.				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib.ch">license@sib.ch</a> ).				
CC	-----				
DR	EMBL: M57860; AAA26100.1; -				
DR	PIR: JQ1013; JQ1013.				
DR	HSSP: P08306; IAR1.				
DR	InterPro: IPR001505; COX2.				
DR	InterPro: IPR002429; Cyl_c_ox_2.				
DR	Pfam: PF00116; COX2; 1.				
DR	ProDom: PD000131; COX2; 1.				
DR	PROSITE: PS00078; COX2; 1.				
KW	Oxidoreductase; Respiratory chain; Electron transport; Transmembrane;				
KW	Copper; Signal.				
KW	Copper: Signal.				
FT	SIGNAL	1	24		POTENTIAL.
FT	CHAIN	25	302		CYTOCHROME C OXIDASE POLYPEPTIDE II.
FT	TRANSMEM	60	80		POTENTIAL.
FT	TRANSMEM	104	124		POTENTIAL.
FT	METAL	217	217		COPPER A (PROBABLE).
FT	METAL	251	251		COPPER A (PROBABLE).

FT METAL 255 255 COPPER A (PROBABLE).  
 FT METAL 259 259 COPPER A (PROBABLE).  
 SQ SEQUENCE 302 AA; 32940 MW; 4D40751A8E4C9AA4 CRC64;

Query Match 71.6%; Score 159; DB 1; Length 302;  
 Best Local Similarity 73.8%; Pred. No. 1,6e-12;  
 Matches 31; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PLEIWTIVPVVILVIFGAFSLPVLFGKQEPFEGIVINVEG 42  
 DB 99 PLEIAWTIVPVILVIAIGAFSLPVLFGKQEPFEGIVINVEG 140

RESULT 2  
 COX2\_PARDE STANDARD; PRT; 298 AA.

AC P08306;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II PRECURSOR (EC 1.9.3.1)  
 DE (CYTOCHROME A3 SUBUNIT 2) (OXIDASE AA(3) SUBUNIT 2).  
 GN CTAC OR CTAB OR COIL.  
 OS Paracoccus denitrificans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Paracoccus.  
 OX NCBI\_TaxID=266;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S1657;  
 RA Raitio M., Jalli T., Saraste M.;  
 RT "Isolation and analysis of the genes for cytochrome c oxidase in  
 RT Paracoccus denitrificans.";  
 RL EMBL J. 6:2825-2833(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC STRAIN-ATCC 13543;  
 RA MEDLINE=88004464; PubMed=2820725;  
 RA Steinuecke P., Steffens G.C.M., Penskus G., Buse G., Ludwig B.;  
 RT "Subunit II of cytochrome c oxidase from Paracoccus denitrificans.  
 RT DNA sequence, gene expression and the protein.";  
 RL Eur. J. Biochem. 167:431-439(1987).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RA MEDLINE=95379947; PubMed=7651515;  
 RA Iwata S., Ostermeier C., Ludwig B., Michel H.;  
 RT "Structure at 2.8-A resolution of cytochrome c oxidase from  
 RT Paracoccus denitrificans.";  
 RL Nature 376:660-669(1995).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 RA MEDLINE=98021406; PubMed=9380672;  
 RA Ostermeier C., Harrenga A., Ernler U., Michel H.;  
 RT "Structure at 2.7-A resolution of the Paracoccus denitrificans two-  
 RT subunit cytochrome c oxidase complexed with an antibody FV  
 RT fragment.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:10547-10553(1997).  
 RN [5]  
 RP FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME  
 RN COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA  
 RN HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND  
 RN CU(B).  
 RN [6]  
 RP CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
 RN 4 FERRICYTOCHROME C.  
 RN [7]  
 RP COFACTOR: TWO COPPER A ATOMS.  
 RN [8]  
 RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
 RN [9]  
 RP SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 RN [10]  
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
 RN between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 RN the European Bioinformatics Institute. There are no restrictions on its  
 RN use by non-profit institutions as long as its content is in no way  
 RN modified and this statement is not removed. usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).

DR EMBL: X05828; CAA29268.1; -  
 DR PIR: S00106; OBP2N.  
 DR PDB: 1AR1; 11-FEB-98.  
 DR InterPro: IPR001505; COX2.  
 DR InterPro: IPR002429; Cyt\_c-ox\_2.  
 DR Pfam: PF00116; COX2; 1.  
 DR PRINTS: PR01166; CYCOXIDASEII.  
 DR ProDom: PD000131; COX2; 1.  
 DR PROSITE: PS00078; COX2; 1.  
 KW Oxidoreductase; Respiratory chain; Electron transport; Transmembrane;  
 KW Copper; Signal; Inner membrane; 3D-structure.  
 FT SIGNAL 1 29  
 FT CHAIN 30 280 CYTOCHROME C OXIDASE POLYPEPTIDE II.  
 FT PROPEP 281 298 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 30 55 PERIPLASMIC.  
 FT TRANSMEM 56 88 CYTOPLASMIC.  
 FT DOMAIN 89 103  
 FT TRANSMEM 104 134  
 FT DOMAIN 135 280  
 FT MOD\_RES 30 30 PERIPLASMIC.  
 FT METAL 210 210 PYROPLASMIC CARBOXYLIC ACID.  
 FT METAL 245 245 COPPER A 1.  
 FT METAL 247 247 COPPER A 1 AND 2.  
 FT METAL 249 249 COPPER A 2.  
 FT METAL 253 253 COPPER A 1 AND 2.  
 FT METAL 256 256 COPPER A 1.  
 FT METAL 256 256 COPPER A 2.  
 FT CONFLICT 160 161 GV -> AF (IN REF. 2).  
 SQ SEQUENCE 298 AA; 32539 MW; 918A64A9E3B93366 CRC64;

Query Match 62.8%; Score 139.5; DB 1; Length 298;  
 Best Local Similarity 60.5%; Pred. No. 4e-10;  
 Matches 26; Conservative 9; Mismatches 7; Indels 1; Gaps 1;

QY 1 PLEIWTIVPVVILVIFGAFSLPVLFGKQEPFEGIVINVEG 42  
 DB 105 PLEIWTIVPVVILVIAIGAFSLPVLFGKQEPFEGIVINVEG 147

RESULT 3  
 COX2\_RICPR STANDARD; PRT; 313 AA.  
 AC Q92DC6;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (CYTOCHROME  
 DE A3 SUBUNIT 2).  
 GN CTAC OR COXB OR RP406.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MADRID E;  
 RA MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sichenitz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RL Nature 396:133-140(1998).  
 RN [2]  
 RP DOMAIN RPE1.  
 RN MEDLINE=20485642; PubMed=11030655;  
 RA Ogata H., Audic S., Barbe V., Artiguenave F., Fournier P.-E.,  
 RA Raoult D., Claverie J.-M.;  
 RT "Selfish DNA in protein-coding genes of Rickettsia.";  
 RL Science 290:347-350(2000).



```

CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME
CC COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA
CC HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND
CC CU(B) (BY SIMILARITY).
CC CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERROCYTOCHROME C.
CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -1- SIMILARITY: CONTAINS 1 REPEAT INSERT DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ235271; CAA1463.1; -.
CC InterPro: IPR001505; COX2.
CC InterPro: IPR002429; Cyl_c-ox_2.
CC Pfam: PF00116; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC DR PRINTS: PR01166; CYCOXIDASEII.
CC DR PRODOM: PD000131; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC KW Oxidoreductase; Respiratory chain; Electron transport; Transmembrane;
CC Heme; Copper; Complete proteome.
CC FT TRANSMEM 39 59 POTENTIAL.
CC FT TRANSMEM 94 114 POTENTIAL.
CC FT TRANSMEM 131 151 POTENTIAL.
CC FT DOMAIN 5 51 REPEAT INSERT.
CC FT METAL 233 233 COPPER A (POTENTIAL).
CC FT METAL 268 268 COPPER A (POTENTIAL).
CC FT METAL 272 272 COPPER A (POTENTIAL).
CC FT METAL 276 276 COPPER A (POTENTIAL).
CC SQ SEQUENCE 313 AA; 36037 MW; 77CDAD8F6A4201E5 CRC64;

Query Match 43.2%; Score 96; DB 1; Length 313;
Best Local Similarity 41.5%; Pred. NO. 9.5e-05;
Matches 17; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVFAGSLPVLFKQOEPFGDIVINVG 42
Db 133 LEIWTIVPILVILAVPSFRILRAEKIPEDLTIKYVG 173

RESULT 4
COX2_APTAU STANDARD; PRT: 199 AA.
ID COX2_APTAU STANDARD; PRT: 199 AA.
AC 003889;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MTCO2 OR COII.
OS Apterix australis (Brown kwl).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Apterygiformes; Apterygidae;
OC Apteryx.
OX NCBI_TaxID=8822;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";
RL (in) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-

```

```

CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.
CC CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76064; AAB61314.1; -.
CC InterPro: IPR001505; COX2.
CC InterPro: IPR002429; Cyl_c-ox_2.
CC Pfam: PF00116; COX2; 1.
CC DR PRODOM: PD000131; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT TRANSMEM 1 1 POTENTIAL.
CC FT TRANSMEM 16 29 POTENTIAL.
CC FT TRANSMEM 30 49 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT DOMAIN 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 128 128 COPPER A (PROBABLE).
CC FT METAL 163 163 COPPER A (PROBABLE).
CC FT METAL 167 167 COPPER A (PROBABLE).
CC FT METAL 171 171 COPPER A (PROBABLE).
CC SQ SEQUENCE 199 AA; 22374 MW; 0AF496B02F92ACB1 CRC64;

Query Match 38.7%; Score 86; DB 1; Length 199;
Best Local Similarity 34.1%; Pred. NO. 0.001;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVFAGSLPVLFKQOEPFGDIVINVG 42
Db 28 VELWVTLIPALVILALPSLIQILYKMEIDEDPDLTKAIG 68

RESULT 5
COX2_CHICK STANDARD; PRT: 227 AA.
ID COX2_CHICK STANDARD; PRT: 227 AA.
AC P18944;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN MTCO2 OR COII.
OS Gallus gallus (chicken).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90230301; PubMed=2329578;
RA Desjardins P., Morais R.;
RT "Sequence and gene organization of the chicken mitochondrial genome.
RT A novel gene order in higher vertebrates.";
RL J. Mol. Biol. 212:599-634(1990)
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.

```



```

RP  SEQUENCE OF 22-54 FROM N.A.
RC  STRAIN=PEKIN BREED; TISSUE=Liver;
RX  MEDLINE=90360426; PubMed=2390786;
RA  Desjardins P., Ramirez V., Morais R.;
RT  "Gene organization of the Pekin duck mitochondrial genome.";
RL  Curr. Genet. 17:515-518(1990).
CC  -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC  CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC  3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC  TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC  A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC  -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC  4 FERRICYTOCHROME C.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC  INNER MEMBRANE.
CC  -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L22476; AAA72039.1; -
DR  EMBL; X68507; CAA48519.1; -
DR  EMBL; X55531; CAA39147.1; -
DR  HSSP; P00404; 10CC.
DR  InterPro; IPR001505; COX2.
DR  InterPro; IPR002429; Cyt_c-ox-2.
DR  Pfam; PF00116; COX2; 1.
DR  PRINTS; PR01166; CYCOXIDASE11.
DR  PRODOM; PD000131; COX2; 1.
DR  PROSITE; PS00078; COX2; 1.
KW  Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
KW  Electron transport; Respiratory chain.
FT  DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT  TRANSSEM 27 47 POTENTIAL.
FT  DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
FT  TRANSSEM 62 81 POTENTIAL.
FT  DOMAIN 82 228 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT  METAL 160 160 COPPER A (PROBABLE).
FT  METAL 195 195 COPPER A (PROBABLE).
FT  METAL 199 199 COPPER A (PROBABLE).
FT  METAL 203 203 COPPER A (PROBABLE).
SQ  SEQUENCE 228 AA; 25553 MW; E1C4FE82797DBE03 CRC64;

Query Match 38.7%; Score 86; DB 1; Length 228;
Best Local Similarity 34.1%; Pred. No. 0.0012;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIYPIVILVIFGAFSLPVLFKQGEPPEDGIYINVG 42
DB 60 VELIMTILPAIVLALPSLQIILYMDEIDEDPLITKAIG 100

RESULT 8
COX2_CAIMO STANDARD; PRT; 228 AA.
AC P50666;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN COI1.
OS Cairina moschata (Muscovy duck).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Cairina.
OX NCBI_TaxID=8855;
RN [1]

```

```

RP  SEQUENCE FROM N.A.
RC  STRAIN=MUSCOVY SHINNEY DUCK; TISSUE=Liver;
RA  Pan Y.F., Lee Y.H.W., Wei Y.H., Chiang A.N.;
RT  "A gene cytochrome C oxidase subunit II in duck mitochondrial DNA:
RT  structural features and sequence evolution.";
RL  Biochem. Int. 30:479-489(1993).
CC  -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC  CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC  3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC  TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC  A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC  -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC  4 FERRICYTOCHROME C.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC  INNER MEMBRANE.
CC  -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X68508; CAA48520.1; -
DR  HSSP; P00404; 10CC.
DR  InterPro; IPR001505; COX2.
DR  InterPro; IPR002429; Cyt_c-ox-2.
DR  Pfam; PF00116; COX2; 1.
DR  PRINTS; PR01166; CYCOXIDASE11.
DR  PRODOM; PD000131; COX2; 1.
DR  PROSITE; PS00078; COX2; 1.
KW  Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
KW  Electron transport; Respiratory chain.
FT  DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT  TRANSSEM 27 47 POTENTIAL.
FT  DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
FT  TRANSSEM 62 81 POTENTIAL.
FT  DOMAIN 82 228 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT  METAL 160 160 COPPER A (PROBABLE).
FT  METAL 195 195 COPPER A (PROBABLE).
FT  METAL 199 199 COPPER A (PROBABLE).
FT  METAL 203 203 COPPER A (PROBABLE).
SQ  SEQUENCE 228 AA; 25523 MW; D34E9E3797DBCBB CRC64;

Query Match 38.7%; Score 86; DB 1; Length 228;
Best Local Similarity 34.1%; Pred. No. 0.0012;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIYPIVILVIFGAFSLPVLFKQGEPPEDGIYINVG 42
DB 60 VELIMTILPAIVLALPSLQIILYMDEIDEDPLITKAIG 100

RESULT 9
COX2_NOTPE STANDARD; PRT; 198 AA.
AC O03892;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MTCO2 OR COI1.
OS Notoprocta perdicaria (Chilean tinamou).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae;
OC Notoprocta.
OX NCBI_TaxID=30464;
RN [1]

```

```

RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RL between molecular and morphological data sets.";
RL (In) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76067; AAB61323.1; -.
CC InterPro: IPR001505; COX2.
CC InterPro: IPR002429; Cyt_c-ox_2.
CC Pfam: PF00116; COX2; 1.
CC ProDom: PD000131; COX2; 1.
CC PROSITE: PS00078; COX2; 1.
CC Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC NON_TER 1
CC TRANSMEM <1 15 POTENTIAL.
CC DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
CC TRANSMEM 30 49 POTENTIAL.
CC DOMAIN 50 198 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC METAL 128 128 COPPER A (PROBABLE).
CC METAL 163 163 COPPER A (PROBABLE).
CC METAL 167 167 COPPER A (PROBABLE).
CC METAL 171 171 COPPER A (PROBABLE).
CC SEQUENCE 198 AA; 22227 MW; 7A27344A157E6A13 CRC64;

Query Match 38.3%; Score 85; DB 1; Length 198;
Best Local Similarity 31.7%; Pred. No. 0.0014;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQDFPEGDIVINVEG 42
DB 28 VELIWTILPAIVLILALPSLIQILYMDIEDPDLTKAIG 68

RESULT 10
COX2_TINMA STANDARD; PRT; 198 AA.
ID COX2_TINMA
AC 003895;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MTCO2 OR COII.
OS Tinamus major (Great tinamou).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae; Tinamus.
OC NCBI_TaxID=30468;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";

```

```

RL (In) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76070; AAB61332.1; -.
CC HSP: P08306; 1A1.
CC InterPro: IPR001505; COX2.
CC InterPro: IPR002429; Cyt_c-ox_2.
CC Pfam: PF00116; COX2; 1.
CC ProDom: PD000131; COX2; 1.
CC PROSITE: PS00078; COX2; 1.
CC Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC NON_TER 1
CC TRANSMEM <1 15 POTENTIAL.
CC DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
CC TRANSMEM 30 49 POTENTIAL.
CC DOMAIN 50 198 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC METAL 128 128 COPPER A (PROBABLE).
CC METAL 163 163 COPPER A (PROBABLE).
CC METAL 167 167 COPPER A (PROBABLE).
CC METAL 171 171 COPPER A (PROBABLE).
CC SEQUENCE 198 AA; 22305 MW; B655564EED39BA89 CRC64;

Query Match 38.3%; Score 85; DB 1; Length 198;
Best Local Similarity 31.7%; Pred. No. 0.0014;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQDFPEGDIVINVEG 42
DB 28 VELIWTILPAIVLILALPSLIQILYMDIEDPDLTKAIG 68

RESULT 11
COX2_CASBE STANDARD; PRT; 199 AA.
ID COX2_CASBE
AC 003890;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MTCO2 OR COII.
OS Casuaris bennetti (Dwarf cassowary).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Casuariiformes; Casuariidae;
OC Casuaris.
OC NCBI_TaxID=30463;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";
RL (In) Mindell D.P. (eds.);

```

```

RT Avian molecular evolution and systematics, pp.1-1. Academic Press,  

CC New York (1997).  

CC  

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  

CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  

CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  

CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINDING COPPER  

CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.  

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  

CC 4 FERRICYTOCHROME C.  

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  

CC INNER MEMBRANE.  

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  

CC  

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  

CC the European Bioinformatics Institute. There are no restrictions on its  

CC use by non-profit institutions as long as its content is in no way  

CC modified and this statement is not removed. Usage by and for commercial  

CC entities requires a license agreement (See http://www.isdb-sib.ch/announce/  

CC or send an email to license@isb.sib.ch).  

CC -----  

CC DR EMBL: U76065; AAB61317.1;  

CC DR Interpro: IPR001505; COX2.  

CC DR Interpro: IPR002429; Cytc_Ox_2.  

CC DR Pfam: PF00116; COX2; 1.  

CC DR ProDom: PD000131; COX2; 1.  

CC DR PROSITE: PS00078; COX2; 1.  

CC KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  

CC Electron transport; Respiratory chain.  

CC FT NON_TER 1 1  

CC FT TRANSMEM <1 15 POTENTIAL.  

CC FT DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).  

CC FT TRANSMEM 30 49 POTENTIAL.  

CC FT DOMAIN 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  

CC FT METAL 128 128 COPPER A (PROBABLE).  

CC FT METAL 163 163 COPPER A (PROBABLE).  

CC FT METAL 167 167 COPPER A (PROBABLE).  

CC FT METAL 171 171 COPPER A (PROBABLE).  

CC SQ SEQUENCE 199 AA; 22378 MW; 62E1D4EB5EF92E8 CRC64;
```

```

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + CC CC CYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).
CC -----
CC DR EMBL; U76066; AAB61320.1; -.
CC DR InterPro; IPR001505; COX2.
CC DR InterPro; IPR002429; Cytochrome_c_ox_2.
CC DR Pfam; PF00116; COX2; 1.
CC DR ProDom; PD000131; COX2; 1.
CC DR PROSITE; PS00078; COX2; 1.
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane; Electron transport; Respiratory chain.
CC FT NON_TER 1 1 POTENTIAL.
CC FT TRANSMEM <1 15 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT DOMAIN 16 29 POTENTIAL.
CC FT TRANSMEM 30 49 POTENTIAL.
CC FT DOMAIN 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 128 128 COPPER A (PROBABLE).
CC FT METAL 163 163 COPPER A (PROBABLE).
CC FT METAL 167 167 COPPER A (PROBABLE).
CC FT METAL 171 171 COPPER A (PROBABLE).
CC SEQUENCE 159 AA; 22336 MW; 62E1DDE537C92E8 CRC64;
CC
QY 2 LEIWTIVPVVLVFIQAFSLPVRKQDFEPFGDIVINWG 42
   :|::|||::|::|::|::|::|::|::|::|::|::|
DB 28 VELWTLIRPAIVLIILLALPSGLQIIYMDEIDEDPLTKAIG 68
-----
RESULT 13
COX2_RHEAM STANDARD; PRT; 199 AA.
AC 003893:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MTCO2 OR COII.
OS Rhea americana (Greater rhea) (Common rhea).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Palaeognathae; Rheiformes; Rheidae; Rhea.
ON NCBI_TaxId=8797;
RX 11
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts between molecular and morphological data sets.";
RL (In) Mindell D.P. (eds.);
AV Avian molecular evolution and systematics, pp.1-1, Academic Press, New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2

```

```

CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTERMEMBRANE PROTEIN, MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76068; AAB61326.1; -
CC DR InterPro: IPR001505; COX2.
CC DR Pfam: PF00116; COX2; 1.
CC DR ProDom: PD000131; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT NON_TER 1 1
CC FT TRANSMEM 1 15 POTENTIAL.
CC FT DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 30 49 POTENTIAL.
CC FT DOMAIN 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 128 128 COPPER A (PROBABLE).
CC FT METAL 163 163 COPPER A (PROBABLE).
CC FT METAL 167 167 COPPER A (PROBABLE).
CC FT METAL 171 171 COPPER A (PROBABLE).
CC SQ SEQUENCE 199 AA; 22265 MW; 58725E39529773B7 CRC64;

Query Match 38.3%; Score 85; DB 1; Length 199;
Best Local Similarity 31.7%; Pred. No. 0.0014;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

Qy 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIYINVEG 42
Db 28 VELIWTILPAIVLILALPSLQILVYMDIEDPDLTKAIG 68

RESULT 14
COX2_STRCA STANDARD; PRT; 229 AA.
AC 021400; 003894;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN MTCO2 OR COII.
OS Struthio camelus (Ostrich).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
OC Struthio.
OX NCBI_TaxID=8801;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=97357422; PubMed=9214748;
RA Hatlid A., Janke A., Arnason U.;
RT "The mtDNA sequence of the ostrich and the divergence between
RT paleognathous and neognathous birds.";
RL Mol. Biol. Evol. 14:754-761(1997).
RN 12
RP SEQUENCE FROM N.A.
RA Sorenson M.D., Dimcheff D.E., Ast J.C., Yuri T., Mindell D.P.;
RT "Primers for a PCR-based approach to complete mitochondrial genome
RT sequencing.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

```

```

RN 13
RP SEQUENCE OF 33-229 FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";
RL (In) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y12025; CA72747.1; -
CC DR EMBL: AF069429; AAD09386.1; -
CC DR EMBL: U76069; AAB61329.1; ALT_TERM.
CC DR HSSP: P08306; IARI.
CC DR InterPro: IPR001505; COX2.
CC DR InterPro: IPR002429; Cyt_c-ox_2.
CC DR Pfam: PF00116; COX2; 1.
CC DR PRINTS: PR01166; CYCOXIDASEII.
CC DR PRODOM: PD000131; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT TRANSMEM 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT DOMAIN 27 47 POTENTIAL.
CC FT TRANSMEM 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 62 81 POTENTIAL.
CC FT DOMAIN 82 229 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 160 160 COPPER A (PROBABLE).
CC FT METAL 195 195 COPPER A (PROBABLE).
CC FT METAL 199 199 COPPER A (PROBABLE).
CC FT METAL 203 203 COPPER A (PROBABLE).
CC SQ SEQUENCE 229 AA; 25699 MW; B9377617BACB14EF CRC64;

Query Match 38.3%; Score 85; DB 1; Length 229;
Best Local Similarity 31.7%; Pred. No. 0.0016;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

Qy 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIYINVEG 42
Db 60 VELIWTILPAIVLILALPSLQILVYMDIEDPDLTKAIG 100

RESULT 15
COX2_CARAU STANDARD; PRT; 230 AA.
AC 078682;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN COXII OR COII.
OS Carassius auratus (Goldfish).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

```

```

CC Cypriniformes: Cyprinidae; Cyprininae; Carassius.
OX MBL_taxid=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A23; TISSUE-Oocyte;
RA Murakami M., Yamashita Y., Fujitani H.;
RT "The complete sequence of mitochondrial genome from a gynogenetic
RT triploid 'gibuna' (Carassius auratus langsdorfi).";
RL Zool. Sci. 15:335-337(1998).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB006953; BAA31241.1; -.
CC DR InterPro; IPR001505; COX2.
CC DR InterPro; IPR002429; CYL_C_OX_2.
CC DR Pfam; PF00116; COX2; 1.
CC DR PRINTS; PR01166; CYCOXIDASE1.
CC DR PRODom; PD000131; COX2; 1.
CC DR PROSITE; PS00078; COX2; 1.
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT TRANSMEM 27 48 PROBABLE.
CC FT DOMAIN 49 62 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 63 82 PROBABLE.
CC FT DOMAIN 83 230 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 161 161 COPPER A (PROBABLE).
CC FT METAL 196 196 COPPER A (PROBABLE).
CC FT METAL 200 200 COPPER A (PROBABLE).
CC FT METAL 204 204 COPPER A (PROBABLE).
CC SEQUENCE 230 AA; 26045 MW; 39A0B54FE31B2377 CRC64;

```

```

Query Match 37.8%; Score 84; DB 1; Length 230;
Best Local Similarity 43.9%; Pred. No. 0.0021;
Matches 18; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

```

```

OY 2 LEIWTIVPVVILVIGAFSLPVLFKQEFPEGDIVINVEG 42
Db 61 IEIWTILPAVILVILALPSLRILYLMDEIDPHLTIKAMG 101

```

Search completed: December 22, 2001, 11:08:07  
Job time: 336 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:07:40 ; Search time 37.53 Seconds  
(without alignments)  
171.489 Million cell updates/sec

Title: US-09-712-768-4  
Perfect score: 222  
Sequence: 1 PLEIWTIVPVVILVIFGA.....LFKQEPPEGDIVNWSRS 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:archaea:17\*\*  
2: SP:bacteria:17\*\*  
3: SP:fungi:17\*\*  
4: SP:human:17\*\*  
5: SP:invertebrate:17\*\*  
6: SP:mammal:17\*\*  
7: SP:mhc:17\*\*  
8: SP:organelle:17\*\*  
9: SP:phage:17\*\*  
10: SP:plant:17\*\*  
11: SP:rodent:17\*\*  
12: SP:virus:17\*\*  
13: SP:vertebrate:17\*\*  
14: SP:unclassified:17\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	139.5	62.8	297	2 Q59663	Q59663 paracoccus
2	100	45.0	299	2 Q9A229	Q9A229 caulobacter
3	99	44.6	285	2 Q59630	Q59630 nitrobacter
4	92	41.4	279	2 Q9RM99	Q9RM99 bradyrhizob
5	90	40.5	228	8 Q9B6T1	Q9B6T1 eudromia el
6	89	40.1	208	8 Q9G6I1	Q9G6I1 sarcophyton
7	89	40.1	229	8 Q37534	Q37534 cryptococcus
8	89	40.1	229	8 Q37534	Q37534 kalharina l
9	88	39.6	247	8 Q9TBV8	Q9TBV8 acropora te
10	87	39.2	208	8 Q9GAL6	Q9GAL6 cryptococcus
11	87	39.2	227	8 Q9XU17	Q9XU17 smilthornis
12	87	39.2	227	8 Q9G5U8	Q9G5U8 lessonia ru
13	87	39.2	227	8 Q9G5U7	Q9G5U7 xolmis pyro
14	87	39.2	229	8 Q9G5U7	Q9G5U7 orchestella
15	87	39.2	230	8 Q9G849	Q9G849 chrysomya n
16	86	38.7	226	8 Q9G2J4	Q9G2J4 phytomyza v
17	86	38.7	226	8 Q9G1H9	Q9G1H9 phytomyza v
18	86	38.7	227	8 Q9Z234	Q9Z234 corvus frug
19	86	38.7	227	8 Q9TBH9	Q9TBH9 chalcites l

20	86	38.7	227	8 Q9TBH8	Q9TBH8 cacomantis
21	86	38.7	227	8 Q9TBH0	Q9TBH0 corythoixoi
22	86	38.7	227	8 Q9TBG9	Q9TBG9 opisthocomu
23	86	38.7	227	8 Q9M091	Q9M091 scytalopus
24	86	38.7	227	8 Q9M090	Q9M090 scelorchillu
25	86	38.7	227	8 Q9M089	Q9M089 scelorchillu
26	86	38.7	227	8 Q9M088	Q9M088 pteroptochu
27	86	38.7	227	8 Q9M087	Q9M087 pteroptochu
28	86	38.7	227	8 Q9M086	Q9M086 pteroptochu
29	86	38.7	227	8 Q9M085	Q9M085 tyrannus me
30	86	38.7	227	8 Q9G5U9	Q9G5U9 agriornis m
31	86	38.7	227	8 Q9G5U6	Q9G5U6 muscigralla
32	86	38.7	227	8 Q9G5U5	Q9G5U5 muscisaxico
33	86	38.7	227	8 Q9G5U4	Q9G5U4 muscisaxico
34	86	38.7	227	8 Q9G2G4	Q9G2G4 muscisaxico
35	86	38.7	227	8 Q9G2G1	Q9G2G1 muscisaxico
36	86	38.7	227	8 Q9G2G0	Q9G2G0 muscisaxico
37	86	38.7	228	8 Q9XK24	Q9XK24 aythya amer
38	86	38.7	228	8 Q9B701	Q9B701 apteryx man
39	85	38.3	96	8 Q9B980	Q9B980 blastophaga
40	85	38.3	227	8 Q9XU05	Q9XU05 falco pereg
41	85	38.3	227	8 Q9TB14	Q9TB14 turnix vari
42	85	38.3	227	8 Q9TB13	Q9TB13 chaetura pe
43	85	38.3	227	8 Q9TB12	Q9TB12 bufo virgln
44	85	38.3	227	8 Q9TB11	Q9TB11 chordelles
45	85	38.3	227	8 Q9TBH7	Q9TBH7 piaya melan

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	297 AA.
Q59663	Q59663			
AC	059663:			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CYTCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) PRECURSOR.			
OS	Paracoccus denitrificans.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;			
OC	Paracoccus.			
OX	NCBI_TaxID=266;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WILD TYPE;			
RC	MEDLINE=88004464; PubMed=2820725;			
RA	Steinbecke P., Steffens G.C.M., Panskus G., Buse G., Ludwig B.;			
RT	"Subunit II of cytochrome c oxidase from Paracoccus denitrificans. DNA			
RT	sequence, gene expression and the protein."			
RL	Eur. J. Biochem. 167:431-439(1987)			
CC	-1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME			
CC	COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA			
CC	HEME A AND CU(A) TO THE BINDING CENTER FORMED BY HEME A3 AND			
CC	CU(B). SUBUNIT II BINDS CU(A) AND CYTOCHROME C.			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4			
CC	FERROCYTOCHROME C.			
CC	-1- COFACTOR: COPPER A AND HEME GROUP.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	EMBL: X05934; CAA29372.1; "			
DR	HSSP: P08306; IARL.			
DR	InterPro: IPR001505; COX2.			
DR	InterPro: IPR002429; Cyt_c-ox_2.			
DR	Pfam: PF00116; COX2.1.			
DR	PRINTS: PR01166; CYCOXIDASEII.			
DR	ProDom: PD000131; COX2.1.			
DR	PROSITE: PS00078; COX2.1.			
KW	Copper; Oxidoreductase; Signal; Transmembrane.			
FT	SIGNAL	1	28	POTENTIAL.
FT	CHAIN	29	297	POTENTIAL.
FT	METAL	211	211	COPPER A (PROBABLE).
FT	METAL	215	215	COPPER A (PROBABLE).

FT METAL 244 244 COPPER A (PROBABLE)  
 FT METAL 248 248 COPPER A (PROBABLE)  
 FT METAL 252 252 COPPER A (PROBABLE)  
 FT METAL 255 255 COPPER A (PROBABLE)  
 SO SEQUENCE 297 AA: 32469 MW: ED3222C47049A1P8 CRC64;

Query Match 62.8%; Score 139.5; DB 2; Length 297;  
 Best Local Similarity 60.5%; Pred. No. 5e-09; Mismatches 7; Indels 1; Gaps 1;  
 Matches 26; Conservative 9; Mismatches 7; Indels 1; Gaps 1;

OY 1 LEIWTIVPVILVIFGAFSLPVLFKQGFPP-EGDIVINEG 42  
 Db 104 PLEIWTIVPVILVIFGAFSLPVLFKQGFPP-EGDIVINEG 146

RESULT 2  
 ID 09A229 PRELIMINARY; PRT; 309 AA.  
 AC 09A229; PRELIMINARY; PRT; 309 AA.  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE CYTOCHROME C OXIDASE, SUBUNIT II.  
 GN CC3407.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 CC Caulobacter.  
 CX NCBI\_TaxID=69394;  
 RX SEQUENCE FROM N.A.  
 RA MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocha I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Urtreback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser J.C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE006001; AAK25369.1; -  
 DR TIGR; CC3407; -  
 KW Complete proteome.  
 SQ SEQUENCE 309 AA: 33248 MW: AF1432A3B136055F CRC64;

Query Match 45.0%; Score 100; DB 2; Length 309;  
 Best Local Similarity 39.0%; Pred. No. 0.00028; Mismatches 16; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQGFPP-EGDIVINEG 42  
 Db 102 IEVIMTVPVILVIFGAFSLPVLFKQGFPP-EGDIVINEG 142

RESULT 3  
 ID 059630 PRELIMINARY; PRT; 285 AA.  
 AC 059630; PRELIMINARY; PRT; 285 AA.  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) PRECURSOR.  
 GN COXB.  
 OS Nitrospira winogradskyi (Nitrospira agilis).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Bradyrhizobium group; Nitrospira; Nitrospira.  
 CX NCBI\_TaxID=913;  
 RX SEQUENCE FROM N.A.  
 RA STRAIN=ATCC 14123;  
 DR Berben G.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -! FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME  
 CC COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA  
 CC HEME A AND CU(A) TO THE BINDING CENTER FORMED BY HEME A3 AND  
 CC CU(B). SUBUNIT II BINDS CU(A) AND CYTOCHROME C.  
 CC -! CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4  
 CC -! COFACTOR: COPPER A AND HEME GROUP.  
 CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -! SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL; X89566; CA61743.1; -  
 DR HSPB; P08306; IAK1.  
 DR InterPro; IPR001505; COX2.  
 DR InterPro; IPR002429; Cyt\_c-ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR ProDom; PD000131; COX2; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 KW Copper; oxidoreductase; signal; transmembrane.  
 FT SIGNAL 1 34  
 FT CHAIN 1 285  
 FT METAL 202 202 COPPER A (PROBABLE)  
 FT METAL 206 206 COPPER A (PROBABLE)  
 FT METAL 235 235 COPPER A (PROBABLE)  
 FT METAL 239 239 COPPER A (PROBABLE)  
 FT METAL 243 243 COPPER A (PROBABLE)  
 FT METAL 246 246 COPPER A (PROBABLE)  
 SQ SEQUENCE 285 AA: 31136 MW: B841475F4CFBD6A0 CRC64;

Query Match 44.6%; Score 99; DB 2; Length 285;  
 Best Local Similarity 40.5%; Pred. No. 0.00034; Mismatches 17; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQGFPP-EGDIVINEG 43  
 Db 105 IEVIMTVPVILVIFGAFSLPVLFKQGFPP-EGDIVINEG 146

RESULT 4  
 ID 09RM99 PRELIMINARY; PRT; 279 AA.  
 AC 09RM99; PRELIMINARY; PRT; 279 AA.  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DE CYTOCHROME C OXIDASE SUBUNIT II PRECURSOR (EC 1.9.3.1) (COXB).  
 GN COXB.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Bradyrhizobium group; Bradyrhizobium.  
 CX NCBI\_TaxID=375;  
 RX SEQUENCE FROM N.A.  
 RA STRAIN=110SPC4;  
 RA Rossmann R., Lofler H., Rossi P., Henneke H.;  
 RT "Factors involved in biogenesis of active cytochrome a3 encoded by  
 RT the COXB gene cluster from Bradyrhizobium japonicum."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA110SPC4;  
 RC MEDLINE=96422470; PubMed=8825087;  
 RA Muller P., Ahrens K., Keller T., Klauke A.;  
 RT "A TnpA insertion within the Bradyrhizobium japonicum sips gene,  
 RT homologous to prokaryotic signal peptidases, results in extensive  
 RT changes in the expression of PBM-specific nodulins of infected soybean  
 RT (Glycine max) cells."  
 RL Mol. Microbiol. 18:831-840(1995).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=USDA110SPC4;  
 RC Mueller P.;

"Extended sequencing of a DNA fragment of B. japonicum adjacent to the  
 RT cox operon."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL: AJ242592; CAB56818.1; -  
 DR EMBL: U33883; AAF78816.1; -  
 DR HSSP: P18400; ICYX.  
 DR InterPro: IPR001505; COX2.  
 DR InterPro: IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam: PF00116; COX2; 1  
 DR PRINTS: PR01166; CYCOXIDASEII.  
 DR PRODOM: PD000131; COX2; 1.  
 DR PROSITE: PS00078; COX2; 1.  
 KW Copper; Oxidoreductase; Signal.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT SEQUENCE 279 AA; 30564 MW; 5FA79C394FA73C8 CRC64;

Query Match 41.4%; Score 92; DB 2; Length 279;  
 Best Local Similarity 40.5%; Pred. No. 0.0023;  
 Matches 17; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 43  
 DB 99 IEVNTIVPVILVIGISVPSRFLFLDLVPRKADLTITATGK 140

RESULT 5  
 ID 09B6T1 PRELIMINARY; PRT; 228 AA.  
 AC 09B6T1.  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE CYTOCHROME C OXIDASE SUBUNIT II.  
 DR Eukarya: Eukarya (Elegant crested-tilamou).  
 OS Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae; Eudromia.  
 OX NCBI\_TaxID=8805;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BREAST MUSCLE;  
 RA MEDLINE=21085666; PubMed=11217857;  
 RA Cooper A., Lalunza-Fox C., Anderson S., Rambaut A., Austin J.,  
 RA Ward R.;  
 RT "Complete mitochondrial genome sequences of two extinct moas clarify  
 RT rattle evolution."  
 RL Nature 409:704-707(2001).  
 DR EMBL: AY016016; AAK08590.1; -  
 KW Mitochondrion.  
 SQ SEQUENCE 228 AA; 25668 MW; 26C2ECCAS813F9E3 CRC64;

Query Match 40.5%; Score 90; DB 8; Length 228;  
 Best Local Similarity 33.3%; Pred. No. 0.0032;  
 Matches 14; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 43  
 DB 60 VELIWTILPAIVLILALPSLQILYMDDEIDPLTKAIGR 101

RESULT 6  
 ID 09G6I9 PRELIMINARY; PRT; 253 AA.  
 AC 09G6I9;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).

GN COII  
 OS Sarcophyton glaucum.  
 OS Mitochondrion.  
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Aleyonaria; Aleyonacea;  
 OC Aleyoniidae; Sarcophyton.  
 OX NCBI\_TaxID=70919;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99065763; PubMed=9847412;  
 RA Beaton M.J., Roger A.J., Cavalier-Smith T.;  
 RT "Sequence analysis of the mitochondrial genome of Sarcophyton  
 RT conserved gene order among octocorals."  
 RL J. Mol. Evol. 47:697-708(1998).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
 CC -1- FERROCYTOCHROME C.  
 CC -1- COFACTOR: COPPER A AND HEME GROUP.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL: AF064823; AAC99650.1; -  
 DR HSSP: P08306; IARL.  
 DR InterPro: IPR001505; COX2.  
 DR InterPro: IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam: PF00116; COX2; 1.  
 DR PRINTS: PR01166; CYCOXIDASEII.  
 DR PRODOM: PD000131; COX2; 1.  
 DR PROSITE: PS00078; COX2; 1.  
 KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;  
 KM Respiratory chain; Transmembrane.  
 FT METAL 186 186 COPPER A (PROBABLE).  
 FT METAL 190 190 COPPER A (PROBABLE).  
 FT METAL 230 230 COPPER A (PROBABLE).  
 FT SEQUENCE 253 AA; 28391 MW; F8B37D4C08939D0F CRC64;

Query Match 40.5%; Score 90; DB 8; Length 253;  
 Best Local Similarity 43.9%; Pred. No. 0.0036;  
 Matches 18; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 42  
 DB 77 IEIWTIIPAILVIFAPPSLKLXMDDEVYRPGVYKAIG 117

RESULT 7  
 ID 09GALL PRELIMINARY; PRT; 208 AA.  
 AC 09GALL;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).  
 GN COII.  
 OS Cryptotermes secundus.  
 OS Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Isoptera;  
 OC Kalotermitidae; Cryptotermitinae; Cryptotermites.  
 OX NCBI\_TaxID=105785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thompson G.J., Lenz M., Miller L., Crozier R.H.;  
 RT "Phylogenetic Analysis and Trait Evolution in Australian Lineages of  
 RT Drywood Termites (Isoptera, Kalotermitidae)."  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-

3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
SIMILARITY).  
-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
FERRICCYTOCHROME C.  
-1- COPRATOR: COPPER A AND HEME GROUP (BY SIMILARITY).  
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL  
INNER MEMBRANE (BY SIMILARITY).  
-1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
DR EMBL: AF189093; AAC17077.1; -  
DR Interpro: IPR001505; COX2.  
DR Interpro: IPR002429; Cyt\_c-ox-2.  
DR Pfam: PF00116; COX2; 1.  
DR PRINTS: PR01166; CYCOXIDASEII.  
DR PRODOM: PD000131; COX2; 1.  
DR PROSITE: PS00078; COX2; 1.  
KW Copper: Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 208 AA; 23745 MW; E787EDAD0D3D91987 CRC64;

Query Match 40.1%; Score 89; DB 8; Length 208;  
Best Local Similarity 41.5%; Pred. No. 0.0039;  
Matches 17; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIFGAFSLPVLFKQEPFEGDIVINVEG 42  
DB 42 LEIWTIVPVVILVIFGAFSLPVLFKQEPFEGDIVINVEG 82

RESULT 8  
Q37534 PRELIMINARY; PRT; 229 AA.  
ID Q37534  
AC Q37534;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-JUN-2001 (CYCOXIDASEII. 17, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).  
OS Katharina tunnicata (black chiton).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Mollusca; Polyplacophora; Neoloricata;  
OC Ischnochitonina; Mopallidae; Katharina.  
OX NCBI\_Taxid=34587;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95129806; Pubmed=782825;  
RA Boore J.L., Brown W.M.;  
RT "Complete DNA sequence of the mitochondrial genome of the black  
chiton, Katharina tunnicata."  
RL Genetics 136:423-443(1994).  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
FERRICCYTOCHROME C.  
CC -1- COPRATOR: COPPER A AND HEME GROUP.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL  
INNER MEMBRANE (BY SIMILARITY).  
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
DR EMBL: U09810; AAC48365.1; -  
DR HSSP: P18400; ICYX.  
DR Interpro: IPR001505; COX2.  
DR Interpro: IPR002429; Cyt\_c-ox-2.  
DR Pfam: PF00116; COX2; 1.  
DR PRINTS: PR01166; CYCOXIDASEII.  
DR PRODOM: PD000131; COX2; 1.  
DR PROSITE: PS00078; COX2; 1.  
KW Copper: Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
SQ SEQUENCE 208 AA; 23745 MW; E787EDAD0D3D91987 CRC64;

KW Copper: Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
FT METAL 163 163 COPPER A (PROBABLE).  
FT METAL 167 167 COPPER A (PROBABLE).  
FT METAL 196 196 COPPER A (PROBABLE).  
FT METAL 200 200 COPPER A (PROBABLE).  
FT METAL 204 204 COPPER A (PROBABLE).  
FT METAL 207 207 COPPER A (PROBABLE).  
SQ SEQUENCE 229 AA; 26079 MW; FA0BAb11965E754E CRC64;

Query Match 40.1%; Score 89; DB 8; Length 229;  
Best Local Similarity 43.9%; Pred. No. 0.0043;  
Matches 18; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIFGAFSLPVLFKQEPFEGDIVINVEG 42  
DB 61 LEIWTIVPVVILVIFGAFSLPVLFKQEPFEGDIVINVEG 101

RESULT 9  
Q9TBV8 PRELIMINARY; PRT; 247 AA.  
ID Q9TBV8  
AC Q9TBV8;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).  
GN COIT.  
OS Acropora tenuis.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;  
OC Astrocenina; Acroporidae; Acropora.  
OX NCBI\_Taxid=70783;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Oppen M.J.H., Hislop N.R., Miller D.J.;  
RT "The mitochondrial genome of the scleractinian coral Acropora tenuis:  
major differences in gene order within the anthozoa subclass  
Zoantharia."  
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
FERRICCYTOCHROME C.  
CC -1- COPRATOR: COPPER A AND HEME GROUP (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL  
INNER MEMBRANE (BY SIMILARITY).  
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
DR EMBL: AF152244; AAD52920.1; -  
DR HSSP: P18400; ICYX.  
DR Interpro: IPR001505; COX2.  
DR Interpro: IPR002429; Cyt\_c-ox-2.  
DR Pfam: PF00116; COX2; 1.  
DR PRINTS: PR01166; CYCOXIDASEII.  
DR PRODOM: PD000131; COX2; 1.  
DR PROSITE: PS00078; COX2; 1.  
KW Copper: Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
SQ SEQUENCE 247 AA; 28024 MW; 9F7D810C4B2F8D06 CRC64;

Query Match 39.6%; Score 88; DB 8; Length 247;  
Best Local Similarity 46.3%; Pred. No. 0.0061;  
Matches 19; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIFGAFSLPVLFKQEPFEGDIVINVEG 42  
DB 77 LEIWTIVPVVILVIFGAFSLPVLFKQEPFEGDIVINVEG 117





AC Q9G849;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).  
 GN COII.  
 OS Chrysomya norrisi.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Oestridae; Calliphoridae; Chrysomya.  
 OX NCBI\_TaxID=142904;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wells J.D., Sperling F.A.;  
 RT "DNA Based Identification of Forensically Important Chrysomyinae  
 (Diptera: Calliphoridae)".  
 RL Forensic Sci. Int. 0:0-0(2001).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BINETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
 CC FERRICYTOCHROME C.  
 CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL: AF295552: AAG34093.1; -  
 DR InterPro: IPR001505: COX2.  
 DR InterPro: IPR002429: CYL\_c-ox\_2.  
 DR Pfam: PF00116: COX2; 1.  
 DR PRINTS: PR01166: CYCOXIDASEII.  
 DR ProDom: PD000131: COX2; 1.  
 DR PROSITE: PS00078: COX2; 1.  
 KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 KW NON\_TER 230  
 FT 230  
 SQ SEQUENCE 230 AA; 26368 MW; 969373C75291F85C CRC64;

Query Match 39.2%; Score 87; DB 8; Length 230;  
 Best Local Similarity 43.9%; Pred. No. 0.0075;  
 Matches 18; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIFGAFSLPYLFKQDEFPESGDIVINVEG 42  
 :||:||||:|:||||| | :|: | | : | |  
 DB 61 IEIWTILPAIILFLFAFSLRLLYLDLEINERSTLKVIG 101

Search completed: December 22, 2001, 11:07:40  
 Job time: 349 sec





Thu Dec 27 08:21:40 2001

us-09-712-768-3.rge

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
December 22, 2001, 09:44:42 ; Search time 2710.12 Seconds  
803.517 Million cell updates/sec

Run on:

Title: US-09-712-768-3  
Perfect score: 132  
Sequence: 1 ccgcgtgaatcgcctgcgcac.....tcaacgtcgaaggtcgttagc 132

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgo\_hum:\*  
31: em\_hgo\_inv:\*  
32: em\_hgo\_rod:\*  
33: em\_hgt\_hum:\*  
34: em\_hgt\_inv:\*  
35: em\_hgt\_rod:\*  
36: em\_hgt\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	132	100.0	132	6	AX113988
2	75.4	57.1	1356	1	M57680
3	48.2	36.5	1653	1	PDOX12
4	48.2	36.5	4890	1	PDOX1
5	47.6	36.1	8121	1	B04242592
6	47.6	35.8	11495	1	B033883
7	47.2	29.8	11188	1	AE006001
8	39.4	28.5	12829	1	NMCOXABC
9	37.6	27.4	1809	1	AE004449
10	36.2	27.4	4500	3	DMU36762
11	36.2	27.4	5206	12	DMU36762
12	36.2	27.4	5206	12	DMU36762
13	35.6	27.0	1260	3	SYNCTCONS
14	35.4	26.8	16537	5	AB047553
15	34.2	25.9	331	9	HS7A8R
16	33.6	25.5	316	9	HS197DIR
17	33.6	25.5	684	9	HSU12690
18	33.6	25.5	684	9	AF346988
19	33.2	25.2	624	3	AF189104
20	33.2	25.0	590	2	AC074564
21	32.6	24.7	317	9	HS13F11R
22	32.6	24.7	197219	2	AC087839
23	32.2	24.4	9810	1	AF018073
24	32.2	24.4	16896	4	MIMRGEN
25	32.2	24.2	182	6	AX063146
26	32.2	24.2	195	6	AX192821
27	32.2	24.2	230	6	AX062955
28	32.2	24.2	236	6	AX192820
29	32.2	24.2	270	6	AX062942
30	32.2	24.2	275	6	AX063008
31	32.2	24.2	275	6	AX193206
32	32.2	24.2	275	6	AX063066
33	32.2	24.2	277	6	AX063116
34	32.2	24.2	281	6	AX062826
35	32.2	24.2	281	6	AX063087
36	32.2	24.2	288	6	AX062978
37	32.2	24.2	294	6	AX062854
38	32.2	24.2	306	6	AX062854
39	32.2	24.2	313	6	AX192819
40	32.2	24.2	313	6	AB019563
41	32.2	24.2	333	9	MTHSCOXII
42	32.2	24.2	586	9	AK026536
43	32.2	24.2	597	9	AF004339
44	32.2	24.2	684	9	HSU12691
45	32.2	24.2	684	9	HSU12691

#### ALIGNMENTS

RESULT	1	AX113988	132 bp	DNA	PAT	01-MAY-2001
LOCUS	AX113988	Sequence	3	from Patent EP1103603.		
DEFINITION	AX113988					
ACCESSION	AX113988.1	GI:13940148				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
Glucanobacter oxydans.						
Bacteria: Proteobacteria; alpha subdivision; Acetobacteraceae;						
Glucanobacter.						
1 (bases 1 to 132)						
Asakura, A., Hoshino, T. and Shinjoh, M.						
Cytochrome c oxidase complex from glucanobacter oxydans						
Patent: EP 1103603-A 3 30-MAY-2001;						
F. HOFFMANN-LA ROCHE AG (CH)						
Location/Qualifiers						
1..132						
/organism="Glucanobacter oxydans"						
/db_xref="taxon:442"						

Query Match	100.08;	Score 132;	DB 6;	Length 132;
Best Local Similarity	100.08;	Pred. No. 4.3e-26;		
Matches 132; Conservative				

## RESULT 2

ORGANISM

JOURNAL of Invertebrate Biology  
Gene 101 (1), 133-137 (1991)  
MEDLINE  
91285423

## AUTHORS

**TITLE** *Direct Submission*  
**JOURNAL** *Submitted (25-JUL-2001)*

REMARK Sequence update by submitter  
COMMENT On Jul 26, 2001 this sequence version was  
FEATURES

Source

gen

CDS

```

/gene="coxII"
349..1260
/gene="coxII"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=1
/product="cytochrome c oxidase subunit II"
/protein_id="AAA26100.3"
/db_xref="GI:15022167"
/transl_table="MMSRRVTTGCATCAAGLAATAAAAOOQSLITIGPQCGGCT
PSASPVATQIHMIDDFITVIAATITFVLLILVAFHEKRRKUPARFQNSGIGT
AMTIVPILVIAIGAFSLPTVFNQDIEPADTVVGTGYOMKWCYCY

```

BASE COUNT	ORIGIN
239 a	ICSPATGDDNKKSEEEVQQLIENGYSRDEFLATDTAMVVPYKTVVVQVGTGAUVIHS WTVPAFPGKQDAPGRLAOLMPFAEREGCTFTGGSCSLGISHAVMPLTVKVSSEAYA AMLEDRGRTLESSTVLPATPACVASE"
447 c	447 c
437 g	437 g
233 t	233 t

Query Match	57.1%	Score	75.4;	DB 1;	Length	1356;
Best Local	Similarity	75.2%				
Matches	94;	Conservative	Pred. No. 1.3e-10;			

	Insertions	Deletions	Mismatches	Indels	Gaps
QY	1	1	31	0	0

Db 643 CCGCTGAGATCGCCTGGAGCAGTCTGTCGGACACTGTTTCTGTTCTGTCATCGTGTGCTC 60

[illegible][illegible]

Accession	Sequence	Position
Uy	121 gaggg	125
Db	763 ACGGG	767

### RESULT 3

PDOX12  
LOCUS  
PDOX12  
1653 bp

DEFINITION	ACROSS	UP	DNA	BCR	21-MAR-1995
P. denitrificans ox12 gene for cytochrome c oxidase subunit 1 (9.3.1).					

```

REVISION      1
VERSION       1
KEYWORDS      X05934
               X05934.1
               GI:45483
               cvfrobrowse

```

SOURCE  
~100,000,000 C. Oxidase subunit II.  
Paracoccus denitrificans.  
Paracoccus denitrificans.  
ORGANISM

REFERENCE

FINANCE	1 (bases 1 to 1653)
AUTHORS	Ludwig, B.
TITLE	Direct Sub...

JOURNAL  
 ACCEPTED FOR SUBMISSION  
 Submitted (31-AUG-1987) Ludwig B., Institut fuer Biochemie  
 Medizinische Universitaet Bonn

2 Luebeck 1 Universität zu Luebeck, Ratzeburger Allee 160, 240

TITLE	STAFF	DATE	REMARKS
steinrucke, P., Steffens, G. C., Panskus, G., Buse, G. and Ludwig, B.			
Subunit II of cytochrome c oxidase from			
DNA sequence			

sequence, gene expression and the protein  
Eur. J. Biochem. 167 (3), 431-439 (1987)  
88004464

Location/Qualifiers	source
1. .1653	

```
/organism="Paracoccus denitrificans"  
/strain="wild type"
```

```
/ad_xref="taxon:266"  
/clone_lib="pUC8"  
/clone="470"
```

sig_peptide	score	/product='sigma'	monte-carlo
CDS	532.615		

```
532. .1425 precursor polypeptide (AA -28 to -1)"
/note="precursor polypeptide (AA -28 to -1)"
```

```

/codon_start=1
/transl_table=11
/protein_id="name"

```

```

    /db_xref="GI:45484"
    /db_xref="SPTREMBL:O60553"

```

```

...:L:Q55063")
/translation="MAIATKRGVAAVMSLGVATMTAVPALADVLGDLPIVIGKPVN
GMNFOPASSPLAHDOQWLDHEVLVYITAVTETGKLVN
...

```

KORVPBCTMOTIHOVAVOYKQWZSWEYUWV  
AFDAMLKEALADAGYSEDEYLLATDNPPVVPVPPKDDPVVA  
PIEVLMTLVPVILLVATGAFNSLPILFRSQEMPNDDLYKAIGHQWWSYEYPDA

mat\_peptide  
616..1422  
EFADASDYLIPASPVKLSAE"

BASE COUNT	/product="mature cytochrome oxidase subunit II (AA 1-269)
270 a	
525 c	
556 d	
302 +	



```

Query Match          36.5%; Score 48.2; DB 1; Length 4890;
Best Local Similarity 63.2%; Pred. No. 0.0034;
Matches 74; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 ccgcgtgaatcgcctgcagacatgtccgcgtgtgcatcgttcacatcgcgtc 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1304 CCATGAGAGTATCTGACACCTGCTCCGCTGCTGATCTGCTGATCGGATCGGCGCTTC 1363
QY 61 tcgcctccgcgtcgtcttcaaacagaagatcccccgagggtgacatcgtcacaac 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1364 TCCTCCGCATCTCTTCCGACGACGAGATGCCGAGATCCGACCTCGCTGATC 1420

RESULT 5
LOCUS      BJA242592          8121 bp      DNA
DEFINITION Bradyrhizobium japonicum coxh, coxa, coxe, coxf, coxg, coxc, shb1
ACCESSION  AU242592
VERSION     AU242592.1 GI:6006409
KEYWORDS    assembly protein; coxa gene; coxh gene; coxc gene; coxe gene; coxf
gene; coxg gene; cytochrome c oxidase subunit I; cytochrome c
oxidase subunit II; cytochrome c oxidase subunit III; heme A
IX farnesyltransferase; shb1 gene; SUR1 homolog; lld gene; lld
homolog;
SOURCE      Bradyrhizobium japonicum.
ORGANISM    Bradyrhizobium japonicum
Bacteria; Proteobacteria;
Bradyrhizobium group; Bradyrhizobium.
REFERENCE   1 (bases 1 to 8121)
AUTHORS     Rossmann, R., Loferer, H., Rossi, P. and Hennecke, H.
TITLE       Factors involved in biogenesis of active cytochrome a3 encoded by
            the coxBACF gene cluster from Bradyrhizobium japonicum
JOURNAL     Unpublished
AUTHORS     Rossmann, R.
TITLE       Direct Submission
JOURNAL     Submitted (20-MAY-1999) Rossmann R., Eidgenossische Technische
            Hochschule Zuerich, Mikrobiologisches Institut, Schmelzbergstrasse
            7, CH-8092 Zuerich, SWITZERLAND
FEATURES    Location/Qualifiers
             1..8121
             /organism="Bradyrhizobium japonicum"
             /strain="110spc4"
             /db_xref="taxon:375"
             /gene="ldd"
             complement(1..440)
             /gene="ldd"
             /function="putative role in regulation of carbon storage"
             /codon_start=1
             /transl_table=11
             /product="ldd homolog"
             /protein_id="CAB56816.1"
             /db_xref="GI:6006410"
             /translation="MTNPPTSILDRALDDQVHRHENVAGLAGADGELFLEYQTE
             ALMFDRKLOATYDTDSGFGIRAKVKKDANGVAVSSVSLPALIRAAADVAAVRGIS
             GFSAPPTNVALYGDNDPDDADEFETKVKLLAEIDAYLRDKD"
             complement(542..1075)
             /note="ORF177"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="CAB56817.1"
             /db_xref="GI:6006411"
             /translation="MGLSRMARPPDSNLNLGLLAALLLALPSLAHAGAVSYVC
             DMOJICDPGCAOECCALIQSVAAEDSRNAGITVIVLKTADOKSRIMRWYVAGVLL
             GEPISLNGIGSGDKLP"
             /gene="coxh"
             /gene="coxh"

gene
CDS
1523..1612
/gene="coxh"
1523..2362
/gene="coxh"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="cytochrome c oxidase subunit II"
/protein_id="CAB56818.1"
/db_xref="GI:6006412"
/translation="MRKSGVGRHLLGLAVAGLLTATGSAFAELGAPAEWTLQO
WTLVPLVINGISVPSFRLLELDVPRKADLTIKATGKNYSSVAPDNGKFEEDSLA
ADDKDRRLGVNDENVVPRNKVIRVQVGTADYTHARALPAFGVCKIDAIPGRINTEWFK
PAKTMFYGQCESELCGKDHAFMPIALRVYEDKFAVSETAKKRFASGCTGYASAG
P70"
2451..4076
/gene="coxh"
2451..4076
/gene="coxh"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="cytochrome c oxidase subunit I"
/protein_id="CAB56819.1"
/db_xref="GI:6006413"
/translation="MATAAAGHAGADHGHDEHAHPTGMRRVYSTNNKIDGTYVLI
FAVIAGVIAAASIAIRAEIYPCVQIHEHHTYNEVSHGLIMEVMAWAMTGF
LSSGHPGAPVAPATLSLHLGASSILGAINFTLTSTVEGPGANGAGATNTP
TYELLISLTPVLAGATIMLTDNRNCTFFAPDGGDVPYLPDHLTPFEGHEPVYIL
PQGMISQIVSTFSKRPVGEGLGMAVMAVIGIGFVMAHMTFTVGSASATQAVPA
ATMVATVAPGVKIFISWIAITMGSIIEFRAPMIVANGFTLFTVGGVTSVLANGVPA
VQETIYVAHFEHVLSIGAVFAIFPGVYWPFKMTGYNITETLAKAHVNTFTGVN
VFPHQETGLSGMPRRYVDYDPAAGMNLVSSVSTISGFLVFLXCVIDAFAKKVP
ADNDMGAGATLTLEMTLPSPPHQFVLPVQ"
4137..5081
/gene="coxh"
4137..5081
/gene="coxh"
/function="prothome IX farnesyltransferase"
/codon_start=1
/transl_table=11
/product="putative heme O synthase"
/protein_id="CAB56820.1"
/db_xref="GI:6006414"
/translation="MSVLDONAVDINPRISAEVGYIALKPRVMSLVIFALVGM
MARGHPVLAITSILCTAAGSAGSALNNALBEDIDAKSPRANRPIRGHITRPA
LPRVMAVAVGTGVEPELFLALTFEPPPHMALALRRSDYRAGIPMLPNVAG
LRTAETLITVLAIVAAAPMALGYDAYGVAVSLITACMLVLAIVNIMRERSOS
5078..5251
/gene="coxh"
5078..5251
/gene="coxh"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="cytochrome c oxidase subunit II"
/protein_id="CAB56821.1"
/db_xref="GI:6006415"
/translation="MIAVADKEPPGIYLTGAOKSKRONSIALALGLVVLFFAV
TWAGPVLVLRPM"
5257..5901
/gene="coxh"
5257..5901
/gene="coxh"
/function="biogenesis of cytochrome c oxidase"
/codon_start=1

```

```

/transl_table=11
/evidence=experimental
/product="putative heme A synthase or assembly protein"
/protein_id="CAB56822.1"
/db_xref="GI:6006416"
/translation="MDHEPTISRDSTRARKVGKGLGRDYLVASTGGVVALWYGAS
YAAPFYNNFCRATGNGTTOVATAAPATGPILAKTIAYFSDNAPGLPMKFBPOSE
IEVNIQVTVFTVNOAARTTGAAYNAVALVGSFYOKINCEFTEDTMAGEK
REMYVYVDPSTIADHENDGSTITLSYTFYVXDPVVKPLASEGDKRGNL"
5981..6877
/gene="coxC"
5981..6877
/gene="coxC"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/product="cytochrome c oxidase subunit III"
/protein_id="CAB56823.1"
/db_xref="GI:6006417"
/translation="MATAGKHNDYHLVDPSPWPAVGLSAPITMAVGATAMNHMSA
APYFGVGVLTVMASWMDVYIEAOKKGDTRVYOLHHRGMILFTASVMEFVA
WMAVFNAALEPPADAVATRDVAVGCGTGAAGACAPGTPRHGTEFDPAHPLLN
TLLTSGCTVTVNHNALLENDROGLKYGLILTVLGALETCVOAEGYSHAAPAGN
VYGTFFMATGPHGFHYLVCTIFLLVCLPRAVAGHTPTPOHIGFEFAWYMHFVDVW
LEFLCTYVWGHGNETWAGAH"
6975..7376
/feature="ORF13"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56824.1"
/db_xref="GI:6006418"
/translation="MNDTAGEPPSSETTVLOSAMRGLACKPCRCGCKLYAGFTLAPS
CDRCGLYAFIDTGDPAIFIMLAGIVVACALIVEVKKYQPYWLAHVAIPLIAT
TLPLRANKSLIIMLQFHKAAPGLVDRAK"
7199..8121
/gene="shb1"
7199..8121
/gene="shb1"
/function="cytochrome a a3 biogenesis"
/codon_start=1
/transl_table=11
/product="SUR1 homolog"
/protein_id="CAB56825.1"
/db_xref="GI:6006419"
/translation="MRADRGQVPAVLAACRAVAADPRHHALATRDVAADRRAI
PPQSGAGPAGPRENNETARKRVAGFALFTLLTAFAVLAQWQLORRRAKHELLAA
LTERLAAPRIALPPAOWAALNPARDERRRFTATPAASPDAAVSSGSAVRKDSG
PCTWAFIPARLISGEMVIVDAGFEVNTMODSVDRVAKKLVYGOPALGTYLRFEP
PGMLTPASRDKRIIMFVRDHYAIAISALGCMVAFYIDLEQPAANIPRPGLVDLH
KDDHQAIVATMRALAGAVLAFGCVWRRRS"
BASE COUNT 1282 a 2639 c 2619 g 1581 t
ORIGIN

Query Match 36.1% Score 47.6; DB 1; Length 8121;
Best Local Similarity 62.7%; Pred. No. 0.0049;
Matches 74; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 ccgcgtgaacatcgctgcagcatgttcggttgatctgtgctcttcacgcgtc 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1814 CTGATCGAGTGGCTGGAGCGTGGTTCGGTCTGATCTCGTGGCATTCGCGTGG 1873

QY 61 tcgctgcggttgcgttcaaacagcagagttcccgagagtgacatcgctcagc 118
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1874 TCGTTCCGCTGCTTCCTCGAGCTCGAGAGCGGAGCCTGACCATCAAGG 1931

RESULT 6
BUJ33883 31495 bp DNA BCT 10-AUG-2000
LOCUS Bradyrhizobium japonicum putative epoxide hydrolase EphB (ephB),
DEFINITION
```

```

ACCESSION U33883
VERSION 1
KEYWORDS SOURCE ORGANISM
REFERENCE
AUTHORS Muller,P., Ahrens,K., Keller,T. and Klaucke,A.
TITLE 1 (bases 1 to 31495)
JOURNAL Bradyrhizobium japonicum
PUBMED 8825087
REFERENCE
AUTHORS Mueller,P.
TITLE Extended sequencing of a DNA fragment of B. japonicum adjacent to the cox operon
JOURNAL Unpublished
AUTHORS Mueller,P.
TITLE 3 (bases 1 to 31495)
JOURNAL Direct Submission
AUTHORS Submitted (15-AUG-1995) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany
REFERENCE 4 (bases 1 to 31495)
AUTHORS Mueller,P.
TITLE Direct Submission
COMMENT Submitted (26-JUN-2000) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany
REMARK Sequence update by submitter
FEATURES
location/Qualifiers
1..31495
/organism="Bradyrhizobium japonicum"
/strain="USD110spc4"
/db_xref="taxon:375"
/clone="C27"
/complement(7..909)
/gene="ephB"
/complement(7..909)
/gene="ephB"
/codon_start=1
/transl_table=11
/product="putative epoxide hydrolase EphB"
/protein_id="AA07892.1"
/db_xref="GI:8708901"
/translation="MRPGTAKQIDAGVSMASRGWPSNGVVAICTAGPYDIIHAYD
VAPDGRGAYRVIIPYLRGYGTHFLSSETPROMARPAAMVVDIIALMDKIDIKKAV
AGDWGARTADIIAALMPDAARMCRSAGYLSSQAGSAPLPSAEILOMWOYQYFAT
BGRGAYEKVTHDPAKLIWKLSPQMKTFDDATPDRSAALDNKHVAITTHINFWRIG
LAKGEAKYFELEKKAAPLISVPTITMEDGANAPHPDPSTYAKKESGRHXYRLTIG
GVGHNLPRAPQAFKAVYDADGA"
/complement(1071..1490)
/gene="ohr"
complement(1071..1490)
```

gene  
CDS

					OslkvdgYtANLLlAMSLlLGTGFEIvFCMLSDKIGRKPllLOGCCLlAAlTFEPPIfK
					MttTSNPNALREKAIfEOGGRGGRSGRGIRRVNGRGYPRLHRAIRHRTTlPRVVOGV
					PDDICGRRERGGOGROGSCLRERGOCSSRRARAGRSLSEDRRHHEDEASVRH
					LPSAAATIGLIflLVITVIVMYGPtlAMLYELPEPTlRTYSMLPYHINMGWFGILL
					PATAItAYSTGDIlAGIMlVFASltVlTGFlLEPTEKDVIKAt"
gene					complement(7350) .7823)
CDS					/gene="orf157"
					complement(7350) .7823)
					/gene="orf157"
					/note="orf157"
					/codon_start=1
					/transl_table=11
					/evidence=noL_experimental
					/product="unknown"
					/protein_id="AAF78797.1"
					/db_xref="GI:8708906"
					/translation="MtsARSVTPLGAltLGASHHYllVAARHGADGARLNGPPDPRTAf
					IAEAREARGAVTLFIRHTSAsITQENADPSVlyDLTTALSRLAPENAGCTHDTPSPD
					DMPKHITKMNONSlQVPVINGRLATGLTWAlTLIEHRARHREVlLOfTGANQ"
gene					complement(7853) .8560)
CDS					/gene="orf135"
					complement(7853) .8560)
					/gene="orf135"
					/note="ORF235"
					/codon_start=1
					/transl_table=11
					/evidence=noL_experimental
					/product="unknown"
					/protein_id="AAF78798.1"
					/db_xref="GI:8708907"
					/translation="MPCGCIADVPAPARRGECEDETSDYVHNIMARCSGACRGDGOQR
					SRCAlRARBRKRPLIOHADDPGRAGSGPPAPARVYLHERBRLGlLRCARRACTARR
					RDRAPRPREAECTAGCTDRXGIDEALPKSDLPKKADPSKPRATODORDAAPRPG
					RRSDDVPYGKLASADRHGGOPRAEGRVEDLSLSWRKPGHlIRGLSHVSRSALPPFA
					VTRlRGSGWLPCAIAH"
gene					8893) .9240
					/gene="unc1"
					/note="orf115"
gene					8893) .11493
					/gene="Fo ATPase operon"
					/note="constitutively expressed"
CDS					8893) .9240
					/gene="unc1"
					/note="ORF115; initial gene of the Fo ATPase operon"
					/codon_start=1
					/transl_table=11
					/evidence=noL_experimental
					/product="unknown"
					/protein_id="AAF78799.1"
					/db_xref="GI:8708908"
					/translation="MEIAIDRPNNVRFPNGSEILISGCPNFEAATSRSKPAQDGDR
					AARASMAIfGRlSELSeLVAGVVGAGlGCGFDRLlSTSPGFIVFLLLGfPAVGvVVV
Query Match	36.1%;	Score 47.6;	DB 1;	Length 31495;	
Best Local Similarity	62.7%;	Pred. No. 0.0047;			
Matches 74;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;	
Db 1	cgcgtgaatacgtctcgtagcaatgttcgggttgatctcgtgtctccacggttcggttc	60			
Db 25236	CTGATCCAGGTGGGCTCGAGCGTCGGTCCGCCTGCTCATCTCGTGGCATCTCGGTCGG	25177			
Qy 61	tgcgtgcggtgcgtctcgtcaaacagcaaggattccccgaagggtgacatcgcatcaacg	118			
Db 25176	TCTGTTCCGCTCTGTTCTCTGAGAGTCGAGCTGCCGAAGCGAGACTGACCATCAAGC	25119			
RESULT 7	AE006001/c	11188 bp	DNA	BCT	28-MAR-2001
LOCUS	Caalobacter crescentus	section 327 of 359 of the complete genome.			
DEFINITION	AE006001 AE005673				
VERSION	AE006001.1	GI:13425117			
KEYWORDS					











gene  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAG03497.1"  
/db\_xref="GI:9945934"  
/translation="MSDAKDVTRRLVGRLLLTVMFAFGALVELDYWCALGIN  
KTAGSAYGGBOVDVGRVREKVOEMTSSNNIDMWEPFRSGDOLVHHGQANQVAFARI  
SDKMTQAOALPISIAPEAAAYHFKTECCPTQOYLQGPESIMPVRFIVDRDLKPDV  
RIVLTAYLFDITARKPPVAGR"  
7028..7915  
/gene="coi1"  
/note="PA0108"  
7028..7915  
/gene="coi1"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome c oxidase, subunit III"  
/protein\_id="AAG03498.1"  
/db\_xref="GI:9945935"  
/translation="MAHSHHYVPAOSKWTITSLVTVTGLOTWPNDTAGHKES  
HGMWIFVGHGIIATYMLPCMGANTIRSRAGLYASQMDKSEFGMGKSEITSEVMFGRK  
FGALVYVHGFAGPWLGECKAGVAMHLMFQYSWPLQTPDPRLTPEPSAYTERK  
LPIINILVTSSTPTVTFAHALKKRNKGPLKAMIALTVLIGIAFIILQAEVYHIN  
ELGITAGAGTISTTFMLTGFCHAVLTALITGLITLRIIRGHFDEHHGFEEASW  
YMHFVDVYIGELFIVYVI"  
complement(7931..8140)  
/gene="PA0109"  
complement(7931..8140)  
/gene="PA0109"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03499.1"  
/db\_xref="GI:9945936"  
/translation="MLKVAIVLLLTATVLSLFSGLFFLYKDGSHSRVYNSITFVAVYL  
AAALVAVANGFSGSLNSHAPMH"  
8136..8950  
/gene="PA0110"  
8136..8950  
/gene="PA0110"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03500.1"  
/db\_xref="GI:9945937"  
/translation="WGEFMQSORLSEAGNRDSRKAGAPRGLPTLVYLGLPVLL  
LDNRKLQADDERKALLASPEARGAEPVSPGOLGGLDPDAVYKRLHGRFEDRRLL  
LTPRRKLNGAGVEVLOPFIQDASGLMLVNRGVAWMTPRSRPTLETDPDVLIDMTW  
VSPHPRRTGYAVOMFALATALLATLVLYGVARAKENHESRSDA"  
8925..9503  
/gene="PA0111"  
8925..9503  
/gene="PA0111"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03501.1"  
/db\_xref="GI:9945938"  
/translation="MSLAIPMERPRRARGQLLAIIGLVGPMILASAMYNKNFW  
PQGSYSYGAALINGCOTPADITQGRSGDEBQQLVTAFCGCEDCQQLVYLARIQNTA  
LGRASAGALIAAGELPADPAPNLVRODPRLORGDLPALYDAKGGGERPLMWVD  
PHGULVYRDADAANGKLLKDQVLLKLSHTG"  
9568..10641  
/gene="PA0112"  
9568..10641  
/gene="PA0112"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03502.1"  
/db\_xref="GI:9945939"  
/translation="WSGATRRKPGFTVAVALATLLAAVVVILGATYRLTHAGLCGPDMPG

[illegible]



```

CDS
/gene="DHR38"
/locus="163..1818"
/codon_start=1
/protein_id="CAA61534.1"
/db_xref="GI:929562"
/db_xref="FLYBASE:FBgn0014859"
/db_xref="SWISS-PROT:P49869"
/translation="MDECCFPLSGGSSASPPAPSOLOQLHTLQSQAQMSHPNNSNS
SNMGNHNSGYNHGHFNALNANSSSSASSLYEINGVSADNFGGQOQOQ
TVLWAGSTQSPKLAITLNRSHAHALJOLNAPNSAASPASADLOAGRLQAP
SOLCAVGDIAACGVKVCCEGCKFRRTVOKSKYVCLADKNCVDRRRRCOF
CRFKCLVAVGWKEVVRVTDLSKGRRLPSKPSPOSPSPISLITLALYSHVDT
PSPCLDYSHYEQSMSEADKVOEYOLTSVDYIKORAEKIPGIFDLPEDELLF
QSASLEFLVRLAYRARIDDPKILFCNGVTLHRTQCLRSRGEMNDIMEFSLNLE
DISAPACIALTLITERHGLRPKRYEOLMITSLRDHYVNAEAKOKHYESRL
LGRPLRSLSVOCLOKRIFFYLKLELDVLPAPALLENMFVTLFP"
misc_feature
827..1023
/gene="DHR38"
/locus="163..1818"
/codon_start=1
/protein_id="CAA61534.1"
/db_xref="GI:929562"
/db_xref="FLYBASE:FBgn0014859"
/db_xref="SWISS-PROT:P49869"
/translation="MDECCFPLSGGSSASPPAPSOLOQLHTLQSQAQMSHPNNSNS
SNMGNHNSGYNHGHFNALNANSSSSASSLYEINGVSADNFGGQOQOQ
TVLWAGSTQSPKLAITLNRSHAHALJOLNAPNSAASPASADLOAGRLQAP
SOLCAVGDIAACGVKVCCEGCKFRRTVOKSKYVCLADKNCVDRRRRCOF
CRFKCLVAVGWKEVVRVTDLSKGRRLPSKPSPOSPSPISLITLALYSHVDT
PSPCLDYSHYEQSMSEADKVOEYOLTSVDYIKORAEKIPGIFDLPEDELLF
QSASLEFLVRLAYRARIDDPKILFCNGVTLHRTQCLRSRGEMNDIMEFSLNLE
DISAPACIALTLITERHGLRPKRYEOLMITSLRDHYVNAEAKOKHYESRL
LGRPLRSLSVOCLOKRIFFYLKLELDVLPAPALLENMFVTLFP"

BASE COUNT      1594 a      1266 c      1061 g      1285 t
ORIGIN

Query Match      27.4%   Score 36.2; DB 3; Length 5206;
Best Local Similarity 57.5%   Pred. No. 6.6;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 20 cgattgtccggttgatctgcttcacgtgctgcgtccggtgctgcttca 79
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 509 CCATTGTCGCGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 450
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 80 aacagcaagagtcgccgaagtgacatcgatcaacgtcgagagtcgacg 132
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 449 CCGTAGAAGTTGCGCTGCGCAACACCATTAATTCTGAGGGAAGTGGC 397
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 13
SYNCTCONS 1260 bp DNA SYN 12-JUN-1993
LOCUS Bacterial cytochrome oxidase subunit II consensus sequence, 5' end.
DEFINITION 107491.1 GI:295264
ACCESSION 107491
VERSION 1
KEYWORDS cytochrome oxidase II.
SOURCE Artificial gene DNA.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1260)
AUTHORS Hernandez-Peter, R., Escamilla, E., Helman-Meneguzzi, F. and
Hernandez, F.
TITLE Computer analysis and modelling of a consensus sequence of
bacterial cytochrome oxidase subunit II
JOURNAL Unpublished (1992)
FEATURES
source
1..1260
/location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
1..1260
/gene="coxIcs"
1..>1260
/EC_number="1.9.3.1"
/notes="This sequence describes the consensus features of
the previously published bacterial cytochrome oxidase
subunit II. Y residues denotes hypervariable nucleotides.;
bacterial consensus sequence; putative"
/codon_start=1
/transl_table=11
/function="terminal oxidase"
/label="COX ICS"
/product="cytochrome oxidase II"
/protein_id="AAA72326.1"

CDS
/gene="DHR38"
/locus="163..1818"
/codon_start=1
/protein_id="CAA61534.1"
/db_xref="GI:929562"
/db_xref="FLYBASE:FBgn0014859"
/db_xref="SWISS-PROT:P49869"
/translation="MDECCFPLSGGSSASPPAPSOLOQLHTLQSQAQMSHPNNSNS
SNMGNHNSGYNHGHFNALNANSSSSASSLYEINGVSADNFGGQOQOQ
TVLWAGSTQSPKLAITLNRSHAHALJOLNAPNSAASPASADLOAGRLQAP
SOLCAVGDIAACGVKVCCEGCKFRRTVOKSKYVCLADKNCVDRRRRCOF
CRFKCLVAVGWKEVVRVTDLSKGRRLPSKPSPOSPSPISLITLALYSHVDT
PSPCLDYSHYEQSMSEADKVOEYOLTSVDYIKORAEKIPGIFDLPEDELLF
QSASLEFLVRLAYRARIDDPKILFCNGVTLHRTQCLRSRGEMNDIMEFSLNLE
DISAPACIALTLITERHGLRPKRYEOLMITSLRDHYVNAEAKOKHYESRL
LGRPLRSLSVOCLOKRIFFYLKLELDVLPAPALLENMFVTLFP"
misc_feature
827..1023
/gene="DHR38"
/locus="163..1818"
/codon_start=1
/protein_id="CAA61534.1"
/db_xref="GI:929562"
/db_xref="FLYBASE:FBgn0014859"
/db_xref="SWISS-PROT:P49869"
/translation="MDECCFPLSGGSSASPPAPSOLOQLHTLQSQAQMSHPNNSNS
SNMGNHNSGYNHGHFNALNANSSSSASSLYEINGVSADNFGGQOQOQ
TVLWAGSTQSPKLAITLNRSHAHALJOLNAPNSAASPASADLOAGRLQAP
SOLCAVGDIAACGVKVCCEGCKFRRTVOKSKYVCLADKNCVDRRRRCOF
CRFKCLVAVGWKEVVRVTDLSKGRRLPSKPSPOSPSPISLITLALYSHVDT
PSPCLDYSHYEQSMSEADKVOEYOLTSVDYIKORAEKIPGIFDLPEDELLF
QSASLEFLVRLAYRARIDDPKILFCNGVTLHRTQCLRSRGEMNDIMEFSLNLE
DISAPACIALTLITERHGLRPKRYEOLMITSLRDHYVNAEAKOKHYESRL
LGRPLRSLSVOCLOKRIFFYLKLELDVLPAPALLENMFVTLFP"

BASE COUNT      1594 a      1266 c      1061 g      1285 t
ORIGIN

Query Match      27.4%   Score 36.2; DB 3; Length 5206;
Best Local Similarity 57.5%   Pred. No. 6.6;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 20 cgattgtccggttgatctgcttcacgtgctgcgtccggtgctgcttca 79
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 509 CCATTGTCGCGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 450
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 80 aacagcaagagtcgccgaagtgacatcgatcaacgtcgagagtcgacg 132
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 449 CCGTAGAAGTTGCGCTGCGCAACACCATTAATTCTGAGGGAAGTGGC 397
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 14
LOCUS AB047553
DEFINITION Plecoglossus altivelis mitochondrial DNA, complete genome.
ACCESSION AB047553
VERSION 1
KEYWORDS AB047553.1 GI:13537245
SOURCE Plecoglossus altivelis mitochondrion DNA.
ORGANISM Plecoglossus altivelis mitochondrion DNA.
REFERENCE 1 (bases 1 to 16537)
AUTHORS Ishiguro, N., Miya, M. and Nishida, M.
TITLE Complete Mitochondrial DNA Sequence of ayu Plecoglossus altivelis
JOURNAL Fish. Sci. 67, 474-481 (2001)
REFERENCE 2 (bases 1 to 16537)
AUTHORS Ishiguro, N., Miya, M. and Nishida, M.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2000) Naoya Ishiguro, Ocean Research Institute,
University of Tokyo, Marine Life Science; Minamidai 1-15-1, Nakano,
Tokyo 164-8639, Japan (E-mail:ishiguro@ori.u-tokyo.ac.jp,
Tel:81-3-5351-6487, Fax:81-3-5351-6488)
FEATURES
source
1..16537
/location/Qualifiers
/organism="Plecoglossus altivelis"
/organelle="mitochondrion"
/db_xref="taxon:61084"
1..68
/product="tRNA-Phe"
69..1013
/gene="12S rRNA"
/notes="putative"
/product="12S ribosomal RNA"
69..1013
/gene="12S rRNA"
1014..1084
/product="tRNA-Val"
1085..2797
/gene="16S rRNA"
/notes="putative"
/product="16S ribosomal RNA"
1085..2797
/gene="16S rRNA"
2798..2871
/notes="codon: UUR"
/product="tRNA-Leu"

```

```

gene      2872..3846
/ gene="ND1"
CDS       2872..3846
/ gene="ND1"
/ transl_table=1
/ codon_start=1
/ product="NADH dehydrogenase subunit 1"
/ protein_id="BAB40601.1"
/ db_xref="GI:13537246"
/ translation="MTTTLTHVINPLAIVPVLLAVAFITLLERVLGYMOLKCPN
LNGVYLQIPADGKLFKEPIRPSSEPLFLAPMLATLTALTMAMPPIYPA
DLNGLIFLALSSLAIVSTLGGMASSKALIGALRAVAQTISEVSLGLISII
LFGSGFLQTEFNVTQESIWILIPAMPIAAMWYIITLAETRAPEDLTEGSELYSCFN
VEYAGGPALFEFLAEIVANILMLNTLSLVLELGAISHIPSELTAACNLMTKAAALSYVE
LWVASYPFRFRYDQIMHLWKNFLPLTALVLMHLALPLAFAGLPOL"
tRNA      3847..3918
/ product="tRNA-Ile"
/ complement(3918..3988)
tRNA      3988..4056
/ product="tRNA-Gln"
/ protein_id="BAB40602.1"
/ db_xref="GI:13537247"
/ translation="MNPYTSLLISALIGTALTFSSSHMLAMWGEINTLAIPLM
TQHSRPSVEAAKRYTITQAAATIMFASTNMTGSDIOLSHPAATITAFMAL
ALKVGLPAVHFWLEVTIGLITGTGILSTNOKIAPALIIOTPALNSVYLGMGLA
STLIGSGGNNOTQIRKILAVSSIAHGMWITVSQVSPSLAIGITVITWMTSAFL
MKATAASLNTLATAMAKTPALIALASLISLGGIPUSGTPKRIILQEMTKGCLP
LANTLAITASLTFYLIRICVAMTLTISPMLGASPSWRFSSNOSLLALSTVGT
GLIPPEFLASFSMA"
tRNA      5112..5183
/ product="tRNA-Trp"
/ complement(5185..5253)
tRNA      /product="tRNA-Ala"
tRNA      /complement(5255..5327)
tRNA      /product="tRNA-Asn"
/ complement(5356..5421)
tRNA      /product="tRNA-Cys"
/ complement(5421..5489)
tRNA      /product="tRNA-Tyr"
gene      5491..7041
/ gene="CO1"
/ gene="CO1"
CDS       5491..7041
/ codon_start=1
/ transl_table=2
/ product="cytochrome oxidase subunit 1"
/ protein_id="BAB40603.1"
/ db_xref="GI:13537248"
/ translation="MAITRWEFSTNKHMDIGTLVITFGAMAGVGTALSLITRAEISOP
GALLDDOIVNYIVTAHAFVMEFVWIMVIMGGFNMWLTLMIGAPIMAFPRANNMF
WLLPSPILLASSGVEAGAGTGVVYPLAGNLAHGAASVDLITFSLHGLISSILG
AINTFTTILNKPRAISOYOTPLFEWAVLITAVALLSLPVLAAIGITMLIDRNLTNT
FTDPAGGDPITIOHLTFWFGFCHPEVYIILIPFGGISHIVAYSGKKEPFGMGMVA
MMAIGLIGFIVAHMHFTYGMNDYPRAFETSATMIIAIPGVKFSMLATLHGSSIAK
ETPLMALIGFIFLTVGGIGIVLANSLSLIDYHDTYVVAHFHYVLSMGAVFALLA
FVHMFPLFSGYTLHSHTWKIHFGLMEFVGNVLTFFQPHLGLAGMPRRYSDPDYVTTM
NTVSSIGSLISLVAVIMPLFLLMEFAAKREVSVELSTVENVLHGCPPIYHFEER
AFVOYVAN"
tRNA      complement(7042..7112)
/ note="codon: UCN"
/ product="tRNA-Ser"
tRNA      7117..7188

```

```

gene      7203..7893
/ gene="CO2"
CDS       7203..7893
/ gene="CO2"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/ transl_table=1
/ codon_start=1
/ transl_except=(pos:7893,aa:TERM)
/ transl_table=2
/ product="cytochrome oxidase subunit 2"
/ protein_id="BAB40604.1"
/ db_xref="GI:13537249"
/ translation="MAHPSQLGPDAAAPMEELHFDHMLMIVLLISTLVLYIIV
WVSTRKLKYLDSQRETEITVYPAVILILALPRLILEMDENDPHTKMGH
QWVSYETDIEDYGFDSYKIPDODILPGOFRLLEDHRRVVESEPIRLVSAEDVL
HSAWPAVGKMDVAPGRLNOTAFIASRPGVFGCCSEIGCANHSPMPTVEAVPLKH
FENMSMLLEDA"
tRNA      7894..7968
/ product="tRNA-Lys"
gene      7970..8137
/ gene="Atpase8"
CDS       7970..8137
/ gene="Atpase8"
/ codon_start=1
/ transl_table=2
/ product="adenosine triphosphatase subunit 8"
/ protein_id="BAB40605.1"
/ db_xref="GI:13537250"
/ translation="MPQLNPAWFAVLVFSMLVFLTVIPKVLGHVESNEPTVQSAK
TNPESNMWPMH"
gene      8128..8810
/ gene="Atpase6"
CDS       8128..8810
/ gene="Atpase6"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/ codon_start=1
/ transl_except=(pos:8809..8810,aa:TERM)
/ transl_table=2
/ product="adenosine triphosphatase subunit 6"
/ protein_id="BAB40606.1"
/ db_xref="GI:13537251"
/ translation="MPLSPDOPMSVPPLGVLALALSLPILPPTFARMLNRLL
TLQGFVNRFTQQLPLVNGVKHAKTILTSMLFLITLNMGLPLPTFTQQLSLN
MGLAVPLMLATVYIIGMRNPTALGHLBPGRVPIPIVLIITISLIPRLALGVR
LTANITAGHLLIOLIAATPAFVLLPLMPTVAILTSVFLITLLEAVAMIGAYPVLL
MSLYLOENV"
tRNA      8811..9595
/ gene="CO3"
CDS       8811..9595
/ gene="CO3"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/ codon_start=1
/ transl_except=(pos:9594..9595,aa:TERM)
/ transl_table=2
/ product="cytochrome oxidase subunit 3"
/ protein_id="BAB40607.1"
/ db_xref="GI:13537252"
/ translation="MAHQAHAHFHMYDSPMPPLTGAVVALMTSGLAIMEFHSVILAV
AGTATLLTMQWMDIVREKTFQGHRTPVOKGLRYGMILFTSEVFETFGFMAFY
HSLAPTEIGECWMPPTGITTLDPEVPLMTAVALLASGYVTVAHNSIMEGRKQTI
HSLTGLILGFEFTPLQLEVEVAFTIADGVSTFPAVAFEGHGLHVIIGSTFLALC
ILROIYHNTSEHHNGFEFAAMVHFVDVVMFLVYSIYMWGS"
tRNA      9596..9666
/ product="tRNA-Gly"
gene      9667..10015
/ gene="ND3"
CDS       9667..10015
/ gene="ND3"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"

```

```
Search completed: December 22, 2001, 09:44:51
Job time: 5717 sec
```

Query Match	26.8%;	Score 35.4;	DB 5;	Length 16537;
Best Local Similarity	55.2%;	Pred. No. 10;		
Matches 69;	Conservative 0;	Mismatches 56;	Totals 0	2

gctggaaatcgtcttgacgatctgttccggttgtatctctgttccttcacgcgtgcgtcttc

Up /582 GATCGAATCATTTGGACTGTCTCCCGCAGTCATCTGATCTCATTCGCCCTCCCTC 7441

63 gctgcggtgtctgttcaaacagcaagagttccccgaggtgtgacatcgtcatcaacgtcga 122

Db 7442 CCTCGCATTTCTTTATCTATGAGCAATCAACGACCTCACCCTTACCATCAAGCTTAT 7501

QY 123 gggtc 127  
1111

Db 7502 GGGCC 7506

## RESULT 15

HS7A8F

LOCUS			
HS7A8R	331 bp	DNA	PRI
H.sapiens Cnc			22-OCT-1995
hs7a8r			

Accession  
GenBank accession number: F03184  
read cp97a8.r1td.  
763104

VERSION	263194.1	GI:1035572
---------	----------	------------

CPG Island; genomic MseI fragment human.

## ORGANISM

**REFERENCE**

1 (bases 1 to 31)

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

**AUTHORS** Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.  
**TITLE** Direct Substitution

Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambs CB3 0ET, UK

REFERENCE  
2 (bases 1 to 331)  
CB10 INQ, England. E-mail contact: humquerry@sanger.ac.uk

AUTHORS	TITLE
Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.	Purification of CpG islands using a methylated

JOURNAL Nat. Genet. 6 (3), 236-244 (1994)  
MEDLINE 94282070

COMMENT  
Vector: pGEM-5Zf(-)  
Clones are null?

clones are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1RO, UK. See <http://www.hgmp.mrc.ac.uk/>.

http://www.hgmp.mrc.ac.uk/ for details  
or contact: biohelp@hgmp.mrc.ac.uk

FEATURES	Location/Qualifiers
source	1. .331

```
/organism="Homo sapiens"
/db xref="taxon:9606"
```

```

/sex="male"

```

```
/dev_stage="adult"  
/tissue_type="blood"
```

```
/clone_11b="CGI-1"  
/clone="7a8"
```

BASE COUNT	93 a	105 c	45 g	82 t	6 others
RIGIN					

Query Match	25.98;	Score 34.2;	DB 9;	Length 331;
Best local similarity				

Local Similarity 56.2%; Pred. No. 25;  
Matches 63; Conservative 0; Mismatches 49; Indels 0

5 tggaaatcgtctgaacgcatltaftccgattatcttatcttgccttttcctt

[illegible]

Thu Dec 27 08:21:40 2001

us-09-712-768-3.rge





Perfect score:	132	US-09-712-768-3
Title:		

Scoring table: IDENTITY\_NUC  
Gapop 10.0 / Gapext 1.0

Searched: 930621 seqs, 428662619 residues

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
post-processing: Minimum Match 0%
                  Maximum Match 100%
```

Listing first 45 summaries

```
Database : N_Geneseq_1101:*
1: /SIDS8/qcndata/geneseq/geneseqn/NA1980.DAT:*
      /geneseq/NA1981.DAT:*
```

- 1: N\_Genseq\_1101: \*
- 2: /SIDS8/gcgdata/genseq/genseqn/NA1980.DAT: \*
- 3: /SIDS8/gcgdata/genseq/genseqn/NA1982.DAT: \*
- 4: /SIDS8/gcgdata/genseq/genseqn/NA1983.DAT: \*
- 5: /SIDS8/gcgdata/genseq/genseqn/NA1984.DAT: \*
- 6: /SIDS8/gcgdata/genseq/genseqn/NA1985.DAT: \*
- 7: /SIDS8/gcgdata/genseq/genseqn/NA1986.DAT: \*
- 8: /SIDS8/gcgdata/genseq/genseqn/NA1987.DAT: \*
- 9: /SIDS8/gcgdata/genseq/genseqn/NA1989.DAT: \*
- 10: /SIDS8/gcgdata/genseq/genseqn/NA1990.DAT: \*
- 11: /SIDS8/gcgdata/genseq/genseqn/NA1991.DAT: \*
- 12: /SIDS8/gcgdata/genseq/genseqn/NA1992.DAT: \*
- 13: /SIDS8/gcgdata/genseq/genseqn/NA1993.DAT: \*
- 14: /SIDS8/gcgdata/genseq/genseqn/NA1994.DAT: \*
- 15: /SIDS8/gcgdata/genseq/genseqn/NA1995.DAT: \*
- 16: /SIDS8/gcgdata/genseq/genseqn/NA1996.DAT: \*
- 17: /SIDS8/gcgdata/genseq/genseqn/NA1997.DAT: \*
- 18: /SIDS8/gcgdata/genseq/genseqn/NA1998.DAT: \*
- 19: /SIDS8/gcgdata/genseq/genseqn/NA1999.DAT: \*
- 20: /SIDS8/gcgdata/genseq/genseqn/NA2000.DAT: \*
- 21: /SIDS8/gcgdata/genseq/genseqn/NA2001.DAT: \*
- 22: /SIDS8/gcgdata/genseq/genseqn/NA2001.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result	Query	Match	Length	DB	ID	Description
--------	-------	-------	--------	----	----	-------------

NO.	Accession	Gene
1	132	AAA91490
2	100.0	AAZ32025
3	32.2	AAZ90082
4	24.4	AAZ32025
5	32.2	AAZ90082
6	24.2	AAZ90082
7	32.2	AAZ90082
8	24.2	AAZ90082
9	32.2	AAZ90082
10	24.2	AAZ90082
11	32.2	AAZ90082

C	12	32	24.2	27.0	22	AAF687100
C	13	32	24.2	27.2	22	AAI29219
C	14	32	24.2	27.5	22	AAF68758
C	15	32	24.2	27.5	22	AAF68808
C	16	32	24.2	27.7	22	AAF68518
C	17	32	24.2	28.1	22	AAF68779
C	18	32	24.2	28.8	22	AAF68670
C	19	32	24.2	28.9	22	AAF68636
C	20	32	24.2	29.4	22	AAF68646
C	21	32	24.2	30.6	22	AAF68546
C	22	32	24.2	31.3	21	AAAF78099
C	23	32	24.2	31.3	21	AAI28837
C	24	32	24.2	31.3	21	AAAL6405
C	25	32	24.2	51.3	22	AAAV87034
C	26	32	24.2	52.0	21	AAAT8262
C	27	32	24.2	53.5	22	AAH35343
C	28	32	24.2	56.3	22	AAH35344
C	29	32	24.2	59.5	21	AAAG9853
C	30	32	24.2	60.8	21	AACT78261
C	31	32	24.2	74.7	22	AAV437121
C	32	32	24.2	78.9	21	AAH35342
C	33	32	24.2	78.9	21	AAH16534
C	34	32	24.2	85.4	16	AAAF05906
C	35	32	24.2	85.4	22	AAF68880
C	36	32	24.2	206.7	22	AAH14061
C	37	32	24.2	156.6	20	AAZ11493
C	38	32	24.2	156.6	20	AAZ11491
C	39	32	24.2	156.6	21	AAZ57203
C	40	32	24.2	165.69	21	AAZ57203
C	41	31.6	23.9	29.3	21	AAF68667
C	42	31.6	23.9	55.9	21	AACT5873
C	43	31.6	23.5	1.73	22	AAF68559
C	44	31	23.5	183	22	AAF68555
C	45	31	23.5	366	22	AAF68607

## RESULT

AAA91490  
ID AAA91490 standard; DNA; 132 BP

AAA91490;

DT 06-AUG-2001 (first entry)

G. oxydans cyclohexane; D-sorbitol; COI complex; COI complex;

oxidative fermentation; electron transfer; aldehyde production; 3WGA production; respiration

KW 2-Keto-L-gluconic acid; zoon production; ds  
anabolic acid production; ketone production; ds

*Glycogenobacter oxydans*

aa	Key	Location/Qualifiers
EH		133

```

/*tag= 3
/product= "COIL"

```

XX  
FN  
EP1103603-A2

XX  
PD 30-MAY-2001.

14-NOV-2000; 2000EP-012410  
PF

PR 17-NOV-1999; 000

PA (NOTE / ...)  
XX XX

XX

DR WPI: 2001-357953/38.  
P-PSDB: AAY97751.

XX New cytochrome c oxidase complex having cytochrome c oxidase activity  
PT from Gluconobacter oxydans DSM 4025, useful in mediating electron  
PT transfer in respiratory chain or producing 2-keto-L-gulononic acid from  
XX L-sorbose or D-sorbitol

PS Claim 14: Page 24; 42pp; English.

CC This sequence encodes a fragment of the Gluconobacter oxydans  
CC cytochrome c oxidase, COI1. The invention relates to a cytochrome c  
CC oxidase COI complex. The COI complex is useful in improving oxidative  
CC fermentation and is an essential component mediating electron transfer  
CC in the respiratory chain. The recombinant microorganism and the  
CC cytochrome c oxidase may be used in the genetic preparation of a  
CC recombinant COI complex and in the production of 2-keto-L-gulononic acid  
CC (2KGA) from L-sorbose or D-sorbitol in a culture medium. The COI is also  
CC useful as a terminal oxidase, oxidizing cytochrome c, an electron  
CC acceptor from an enzyme belonging to dehydrogenase for the production of  
CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,  
CC especially the production of 2KGA from L-sorbose or D-sorbitol.

XX Sequence 132 BP; 22 A; 34 C; 40 G; 36 T; 0 other;

Query Match  
Best Local Similarity 100.0%; Score 132; DB 22; Length 132;  
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgtggaatcgtcgtggaagatgttcgcgtgtgtatcgtgtcttcacgtgcgttc 60  
Db 1 ccgtggaatcgtcgtggaagatgttcgcgtgtgtatcgtgtcttcacgtgcgttc 60  
QY 61 tccgtcgcgtgtcgttcacacgaagagttccgcggaggtgcacatcgttcacgttc 60  
Db 61 tccgtcgcgtgtcgttcacacgaagagttccgcggaggtgcacatcgttcacgttc 120  
QY 121 gaggtgcgtgac 132  
Db 121 gaggtgcgtgac 132

RESULT 2  
ID AA232025 standard; DNA; 9810 BP.  
XX AC AA232025;

DT 10-JAN-2000 (first entry)

DE Human METH1 related EST AF018073.

KW Human: METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;  
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;  
KW angiogenesis inhibitor; abnormal wound healing; inflammation;  
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;  
KW diabetic retinopathy; macula degeneration; haemangioma; detection;  
KW arterial-venous malformation; immune deficiency; ss.

OS Homo sapiens.

XX WO9337660-A1.

XX 29-JUL-1999.

XX 22-JAN-1999; 99MO-US01313.

XX 23-JAN-1998; 98US-0072298.

XX 28-AUG-1998; 98US-0098539.

PA (IRUE/) IRUELA-ARISPE L.

PA (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

XX IrueLA-Arispe L, Hastings GA, Ruben SM;

DR WPI: 1999-590684/50.

XX New isolated metalloprotease thrombospondin polypeptides, useful for  
PT treating hyperproliferative disorders, cancers or autoimmune disorders  
PT -  
XX

PS Disclosure: Page 353-359; 457pp; English.

CC AA232000 and AA232001 encode, and AA232002 represent, human  
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2  
CC respectively. METH1 and METH2 have been found to be potent inhibitors of  
CC angiogenesis both in vitro and in vivo. They can be used for treating  
CC cancer and other disorders related to angiogenesis including abnormal  
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,  
CC macula degeneration, haemangiomas, and arterial-venous malformations.  
CC They may be useful in treating deficiencies or disorders of the immune  
CC system, by activating or inhibiting the proliferation, differentiation,  
CC or mobilisation (chemotaxis) of immune cells. The etiology of these  
CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or  
CC toxins), or infectious. They can also be used to treat inflammatory or  
CC conditions, both chronic and acute conditions. The products can also be  
CC used for detection and diagnosis. AA232002 to AA232080, and AA232080 to  
CC AA232080 represent sequences given in the exemplification of the present  
XX invention.

XX Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;

Query Match  
Best Local Similarity 24.4%; Score 32.2; DB 20; Length 9810;  
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 3 gctggaatcgtcgtggaagatgttcgcgtgtgtatcgtgtcttcacgtgcgttc 62  
Db 4608 gccggagatcgtcgtggaagatgttcgcgtgtgtatcgtgtcttcacgtgcgttc 62  
QY 63 gctggaatcgtcgtggaagatgttcgcgtgtgtatcgtgtcttcacgtgcgttc 4667  
Db 4668 gacgtcgcagctgacacacctcgcggcgcccc 4700

RESULT 3  
ID AAC90082 standard; DNA; 9810 BP.  
XX AC AAC90082;

DT 19-MAR-2001 (first entry)

DE AF018073 cDNA clone.

KW METH: metalloprotease; thrombospondin; angiogenesis inhibition;  
KW cancer therapy; benign tumour; ocular angiogenic disease;  
KW rheumatoid arthritis; psoriasis; wound healing; nonunion fracture;  
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;  
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;  
KW coronary collateral; cerebral collateral; arteriovenous malformation;  
KW ischemic limb angiogenesis; Osler-Weber syndrome; wound granulation;  
KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;  
KW angiodioma; fibromuscular dysplasia; expressed sequence tag;  
KW Crohn's disease; atherosclerosis; birth control; ss.

XX Unidentified.

XX WO200071577-A1.

XX 30-NOV-2000.

XX 25-MAY-2000; 2000WO-US14462.  
 PF  
 XX 25-MAY-1999; 99US-0318208.  
 PR 20-JUL-1999; 99US-0144882.  
 PR 10-AUG-1999; 99US-0147823.  
 PR 13-AUG-1999; 99US-0373658.  
 PR 22-DEC-1999; 99US-0171503.  
 PR 22-FEB-2000; 2000US-0183792.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA (IRUE/) IRUELA-ARISPE L.  
 PA (HAST/) HASTINGS G. A.  
 PA (RUBE/) RUBEN S. M.  
 PA (JONA/) JONAK Z. L.  
 PA (TRUL/) TRULLI S. H.  
 PA (FORN/) FORMWALD J. A.  
 PA (TERR/) TERRETT J. A.  
 XX  
 PI IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;  
 PI Formwald JA, Terrett JA;  
 DR WPI: 2001-025136/03.  
 XX  
 PT METH1 and METH2 polynucleotides and encoded polypeptides, used to  
 PT inhibit angiogenesis in the treatment of disorders such as cancer,  
 PT rheumatoid arthritis and psoriasis -  
 PS Claim 7: Pages 653-659; 768pp: English.  
 XX  
 CC The present invention relates to human METH1 and METH2, (ME for  
 CC Metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).  
 CC The present sequence is an expressed sequence tag (EST) for METH. METH  
 CC can be used for inhibiting angiogenesis in an individual, and for  
 CC treating cancer, benign tumours, an ocular angiogenic disease,  
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,  
 CC vasculogenesis, granulomas, hypertrophic scars, nonunion fractures,  
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,  
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,  
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque  
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,  
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or  
 CC atherosclerosis. METH can also be used in birth control. METH can also  
 CC be used in diagnostic methods for the prognosis of cancer.  
 XX  
 SQ Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;  
 Query Match 24.4%; Score 32.2; DB 22; Length 9810;  
 Best Local Similarity 59.1%; Pred. No. 1.7;  
 Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
 OY 3 gctggaaatcgtcgtgacatgttcggtgtgtatctgtgtcattcgtgcttc 62  
 Db 4608 ggcgggacgtcgtcgtgacgtcttcgtgaactgtcgtcgtgaacgagcctctg 4667  
 OY 63 gctgcggtgtgtgtcacaacagcaagattcc 95  
 Db 4668 gacgctgacgtgaccaccctcgggcgcccc 4700  
 RESULT 4  
 AAF68838  
 ID AAF68838 standard; CDNA: 182 BP.  
 XX AAF68838:  
 AC 12-APR-2001 (first entry)  
 XX  
 DT Human lung tumour protein related nucleotide sequence SEQ ID NO:773.  
 XX

KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
 KW cytostatic; antisense inhibition; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200100828-A2.  
 PD 04-JAN-2001.  
 XX  
 PF 30-JUN-2000; 2000WO-US18061.  
 XX  
 PR 30-JUN-1999; 99US-0346492.  
 PR 15-OCT-1999; 99US-0419356.  
 PR 17-DEC-1999; 99US-0466867.  
 PR 30-DEC-1999; 99US-0476300.  
 PR 06-MAR-2000; 2000US-0519642.  
 PR 22-MAR-2000; 2000US-0533077.  
 PR 10-APR-2000; 2000US-0546259.  
 PR 27-APR-2000; 2000US-0560406.  
 PR 05-JUN-2000; 2000US-0589184.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
 PI Retter MW, Mannion J;  
 XX  
 DR WPI: 2001-071488/08.  
 XX  
 PT Lung tumor-associated proteins and the nucleic acids that encode them,  
 PT useful for preventing, diagnosing and treating lung cancer -  
 PS Claim 4: Page 402; 436pp: English.  
 XX  
 CC The present invention describes immunogenic portions of lung tumour-  
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
 CC (I) have cytostatic activity and can be used in gene therapy, antisense  
 CC inhibition and in vaccines. The NAs and the lung tumour-associated  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with their inappropriate expression,  
 CC especially lung cancers. For example, the NAs may be administered to  
 CC treat diseases by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of the protein by expressing inactive proteins  
 CC or to supplement the activity of the protein production of (I). Additionally,  
 CC NAs may be used to produce the lung-tumour associated protein, according  
 CC to standard recombinant DNA methodology. Conversely, antisense NA  
 CC molecules may be administered to down regulate protein expression by  
 CC binding with the cells own genes and preventing their expression. The NA  
 CC and complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar NA sequences in  
 CC samples, and hence which patients may be in need of treatment for lung  
 CC cancer. The (I) may be used as antigens in the production of antibodies  
 CC and in assays to identify modulators (agonists and antagonists) of the  
 CC expression and activity of the protein. AAF68083 to AAF68878 and  
 CC AAB76848 to AAB76878 represent human lung tumour protein related  
 CC nucleotide and protein sequences which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 182 BP; 55 A; 65 C; 20 G; 42 T; 0 other;  
 Query Match 24.2%; Score 32; DB 22; Length 182;  
 Best Local Similarity 55.4%; Pred. No. 0.63;  
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 OY 5 tggaaatcgtcgtgacatgttcggtgtgtatctgtgtcattcgtgcttcgc 64  
 Db 62 tagaacgtcgtgaactacccggtccatcatcctagctccatccctccatcc 121  
 OY 65 tgcggtgtgtcacaacagcaaggttcccgagggtgaactcgtcatca 116  
 Db 122 tagcatccttaccataacagcaggtgaactcctccctaccatca 173

RESULT 5  
AAAT8101  
ID AAA78101 standard; cDNA: 195 BP.  
XX  
AC AAA78101;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE CDNA encoding human colon tumour polypeptide, SEQ ID NO:388.  
XX  
KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;  
XX Immunotherapy; diagnosis; progression; ss.  
XX Homo sapiens.  
OS  
PN WO200037643-A2.  
XX  
PD 29-JUN-2000.  
XX  
PF 23-DEC-1999; 99WO-US30909.  
XX  
PR 23-DEC-1998; 98US-0221298.  
PR 02-JUL-1999; 99US-0347496.  
PR 02-SEP-1999; 99US-0401064.  
PR 19-NOV-1999; 99US-044242.  
PR 02-DEC-1999; 99US-0454150.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;  
PI Wang T, Yujin Y;  
PI WPI: 2000-442671/38.  
XX  
PT New colon tumor polypeptides used to inhibit the development of cancer,  
PT especially colon cancer, and for diagnosing and monitoring the  
PT progression of the cancer -  
PS  
PS Claim 1; Page 203; 229pp; English.

Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or  
portions of proteins which are associated with human colon tumours.  
The invention also specifically discloses 8 human colon tumour proteins  
(AA81897-B11904). The nucleic acids, the polypeptides they encode, and  
antigen presenting cells (APCs, preferably dendritic cells) expressing  
such polypeptides may be used in vaccines that target tumour cells,  
especially colon tumour cells, thereby inhibiting the development of  
cancer. T-cells specific for the polypeptide expressed by the APC are  
used to remove tumour cells from biological samples, especially blood or  
fractions thereof. The sample or the isolated T-cells specific for the  
polypeptide can then be used to inhibit cancer development. CD4+ and/or  
CD8+ T-cells from a patient may be incubated with a polypeptide or  
nucleic acid of the invention, or an APC expressing such a polypeptide,  
to cause the proliferation of specific T-cells. The T-cells can be  
cloned and then administered back to the patient to inhibit cancer  
development. Nucleic acids encoding the polypeptides and antibodies  
against the polypeptides may be used to determine the expression level  
of a tumour protein of the invention, and therefore to determine whether  
cancer cells are present. Such diagnostic methods may also be used to  
monitor the progression of a cancer by repeating the processes at time  
intervals, and comparing the current result to previous results. The  
CC present sequence represents a cDNA encoding a human colon tumour  
CC polypeptide.

Sequence 195 BP: 59 A; 70 C; 22 G; 44 T; 0 other;

Query Match 24.2%; Score 32; DB 21; Length 195;  
Best Local Similarity 55.4%; Pred. No. 0.65;  
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtctgacgatgttcggtgtgattcgtcttcacatcgtgcttcctgcg 64

Db 64 tagaaaccgtctgacctaccctccgcatcaccatccatgctccatccatccatcc 123  
OY 65 tggcgggtgctgttcaaacagaagatcccccgggtgacatgcatataa 116  
Db 124 taagcatccttaccatacagacgaggtcaacgataccctccctaccatcaa 175

RESULT 6  
AAI28839  
ID AAI28839 standard; cDNA: 195 BP.  
XX  
AC AAI28839;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Colon tumour related determined cDNA sequence for clone 31998.  
XX  
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
KW Immunogenic; gene therapy; vaccine; colonic cancer; ss.  
XX  
OS Homo sapiens.  
PN WO200149716-A2.  
XX  
PD 12-JUL-2001.  
XX  
PF 29-DEC-2000; 2000WO-US35596.  
XX  
PR 30-DEC-1999; 99US-0476296.  
PR 10-JAN-2000; 2000US-0480321.  
PR 15-FEB-2000; 2000US-0504629.  
PR 06-MAR-2000; 2000US-0519444.  
PR 19-MAY-2000; 2000US-0575251.  
PR 29-JUN-2000; 2000US-0609448.  
PR 28-AUG-2000; 2000US-0649811.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;  
PI King GE, Wang T, Jiang Y;  
PI WPI: 2001-441847/47.  
XX  
PT Colon tumor associated proteins and nucleic acids useful for the  
PT prevention, diagnosis and treatment of colonic cancer -  
PS  
PS Claim 2; Page 240; 472pp; English.

The present invention describes colon tumour associated proteins (I) and  
the polynucleotides (II) that encode them. (I) have cytostatic activity.  
(I) and (II) can be used in gene therapy and vaccine production. (I) and  
(II) may be used in the prevention, diagnosis and treatment of diseases  
associated with inappropriate colon tumour associated protein (TCAP)  
expression, such as colonic cancer. For example, (I) and (II) may be  
used to treat disorders associated with decreased expression by  
rectifying mutations or deletions in a patient's genome that affect the  
activity of TCAPs by expressing inactive proteins or to supplement the  
patients own production of them. Additionally, (II) may be used to  
produce the TCAP proteins, by inserting the nucleic acids into a host  
cell culturing the cell to express the protein. (II) and its  
complementary sequences may also be used as DNA probes in diagnostic  
polymerase chain reaction (PCR) and hybridisation assays to detect and  
quantitate the presence of similar nucleic acids in samples, and  
therefore which patients may be in need of restorative therapy. (I) may  
also be used as antigens in the production of antibodies against TCAPs  
and in assays to identify modulators of TCAP expression and activity.  
Anti-(I) antibodies and antagonists may also be used to down regulate  
TCAP expression and activity. The anti-(I) antibodies may also be used  
as diagnostic agents for detecting the presence of TCAPs in samples  
(e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512  
and AAI24494 to AAI24523 represent nucleotide and amino acid sequences  
given in the exemplification of the present invention.









PR 30-DEC-1999; 99US-0476300.  
 PR 06-MAR-2000; 2000US-0519642.  
 PR 22-MAR-2000; 2000US-053077.  
 PR 10-APR-2000; 2000US-0546259.  
 PR 27-APR-2000; 2000US-0560406.  
 PR 05-JUN-2000; 2000US-0589184.  
 PA (CORI-) CORIXA CORP.  
 PI Wang T, Bangur GS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
 PI Retter MW, Mannion J;  
 XX WPI; 2001-071488/08.  
 PT Lung tumor-associated proteins and the nucleic acids that encode them,  
 PT useful for preventing, diagnosing and treating lung cancer -  
 XX  
 PS Claim 4; Page 359; 436pp; English.  
 CC The present invention describes immunogenic portions of lung tumour-  
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
 CC (I) have cytostatic activity and can be used in gene therapy, antisense  
 CC inhibition and in vaccines. The NAs and the lung tumour-associated  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with their inappropriate expression,  
 CC especially lung cancers. For example, the NAs may be administered to  
 CC treat diseases by rectifying mutations or deletions in a patient's genome  
 CC or to supplement the activity of the protein by expressing inactive proteins  
 CC to standard recombinant DNA methodology. Conversely, antisense NA  
 CC molecules may be administered to down regulate protein expression by  
 CC binding with the cells own genes and preventing their expression. The NA  
 CC and complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar NA sequences in  
 CC samples, and hence which patients may be in need of treatment for lung  
 CC cancer. The (I) may be used as antigens in the production of antibodies  
 CC and in assays to identify modulators (agonists and antagonists) of the  
 CC expression and activity of the protein. AAF68083 to AAF68878 and  
 CC AAB76848 to AAB76878 represent human lung tumour protein related  
 CC nucleotide and protein sequences which are used in the exemplification  
 CC of the present invention.  
 XX  
 SO Sequence 270 BP; 74 A; 30 C; 93 G; 73 T; 0 other;  
 Query Match 24.2%; Score 32; DB 22; Length 270;  
 Best Local Similarity 55.4%; Pred. No. 0.71;  
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 5 tggaaatcgtctggacgattgttcggttgatcttgatcgcggtgcttcgcg 64  
 Db 121 TAGAAGCCTGTGAACATCTCTGCCGCGCATCTCTAGTCTCATGGCCCTCCATCCG 62  
 QY 65 tggcggctgctgttcaaacagcagaggttcccgaggggtgacatcgtcatca 116  
 Db 61 TACGCATCTTTACATACAGACGAGGTCAACGATCCCTCCCTTACCATCAA 10  
 RESULT 13  
 AAI29219/c  
 ID AAI29219 standard; cDNA; 272 BP.  
 AC AAI29219;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Colon tumour related determined cDNA sequence for clone R0094.C02.  
 XX  
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 KW immunogenic; gene therapy; vaccine; colonic cancer; ss.  
 XX  
 OS Homo sapiens.

XX  
 PN W0200149716-A2.  
 XX 12-JUL-2001.  
 PD 29-DEC-2000; 2000WO-US355596.  
 XX  
 PR 30-DEC-1999; 99US-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504639.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 PA (CORI-) CORIXA CORP.  
 PI Xu J, Lodes MJ, Secret H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jiang Y;  
 XX WPI; 2001-441847/47.  
 DR  
 PT Colon tumor associated proteins and nucleic acids useful for the  
 PT prevention, diagnosis and treatment of colonic cancer -  
 XX  
 PS Claim 2; Page 352; 472pp; English.  
 CC The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (I) and (II) may be  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC patients own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (I) may  
 CC also be used as antigens in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AAI28460 to AAI29512  
 CC given in the exemplification of the present invention.  
 XX  
 SO Sequence 272 BP; 74 A; 30 C; 93 G; 75 T; 0 other;  
 Query Match 24.2%; Score 32; DB 22; Length 272;  
 Best Local Similarity 55.4%; Pred. No. 0.71;  
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 5 tggaaatcgtctggacgattgttcggttgatcttgatcgcggtgcttcgcg 64  
 Db 121 TAGAAGCCTGTGAACATCTCTGCCGCGCATCTCTAGTCTCATGGCCCTCCATCCG 62  
 QY 65 tggcggctgctgttcaaacagcagaggttcccgaggggtgacatcgtcatca 116  
 Db 61 TACGCATCTTTACATACAGACGAGGTCAACGATCCCTCCCTTACCATCAA 10  
 RESULT 14  
 AAF68758/c  
 ID AAF68758 standard; cDNA; 275 BP.  
 AC AAF68758;  
 XX







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 09:36:14 ; Search time 113.93 Seconds  
(Without alignments)  
262.399 Million cell updates/sec

Title: US-09-712-768-3  
Perfect score: 132  
Sequence: 1 ccgcgtggaatcgtctgac.....tcaacgtcagaggtcgtacg 132

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA : \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq : \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq : \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq : \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq : \*  
5: /cgn2\_6/ptodata/2/ina/PCUTS.COMB.seq : \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq : \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	24.2	251	4	US-09-328-111-529
2	32	24.2	353	4	US-09-385-982-410
3	32	24.2	754	1	US-08-219-842-2
4	32	24.2	754	1	US-08-451-096-2
5	32	24.2	789	4	US-09-385-982-539
6	32	24.2	854	4	US-08-413-740A-2
7	32	24.2	854	5	PCT-US95-04063-2
8	32	24.2	16569	4	US-09-097-889-1
9	31.2	23.6	6744	4	US-09-097-889-1
10	29.4	22.3	591	2	US-09-156-979-1
11	28	21.2	30001	1	US-08-125-468-1
12	28	21.2	30001	2	US-08-474-933-1
13	27.8	21.1	9551	1	US-08-056-200-93
14	27.8	21.1	9551	2	US-08-056-200-93
15	27.6	20.9	1600	2	US-08-776-210-3
16	27.6	20.8	571	2	US-08-803-899-1
17	27.2	20.6	783	2	US-08-505-218-9
18	27.2	20.6	1067	2	US-08-505-218-5
19	27	20.5	840	2	US-08-924-759-13
20	27	20.5	840	2	US-09-248-335-13
21	27	20.5	441529	4	US-09-103-840A-1
22	26.8	20.3	1557	5	PCT-US91-01327-11
23	26.8	20.3	1557	6	PCT-US91-01327-11
24	26.6	20.2	971	3	US-09-248-335-65
25	26.6	20.2	11561	1	US-08-450-332-1
26	26.6	20.2	11561	2	US-08-637-640-1
27	26.6	20.2	11561	4	US-09-004-406C-1

28	26.4	20.0	28804	2	US-08-592-874-1	Sequence 1, Appl
29	26.4	20.0	28804	3	US-09-096-942-2	Sequence 2, Appl
30	26.4	20.0	28804	3	US-09-096-867-2	Sequence 2, Appl
31	26	19.7	726	3	US-08-660-645A-11	Sequence 11, Appl
32	26	19.7	726	3	US-09-298-718-11	Sequence 11, Appl
33	26	19.7	726	4	US-09-546-969-11	Sequence 28, Appl
34	26	19.7	726	4	US-08-980-832-28	Sequence 1, Appl
35	26	19.7	1113	4	US-09-172-353-1	Sequence 12, Appl
36	26	19.7	1536	4	US-09-352-990-17	Sequence 32, Appl
37	26	19.7	2632	2	US-08-899-324-13	Sequence 10, Appl
38	26	19.7	2632	2	US-08-329-892B-32	Sequence 32, Appl
39	26	19.7	3023	4	US-09-593-989-10	Sequence 610, App
40	25.8	19.5	254	4	US-09-328-111-610	Sequence 3, Appl
41	25.8	19.5	1995	2	US-08-425-069-3	Sequence 3, Appl
42	25.8	19.5	1995	2	US-08-317-844B-3	Sequence 11, Appl
43	25.6	19.4	1103	3	US-08-617-860B-5	Sequence 5, Appl
44	25.6	19.4	2943	1	US-08-042-747A-7	Sequence 7, Appl
45	25.4	19.2	710	4	US-08-998-416-817	Sequence 817, App

## ALIGNMENTS

RESULT 1  
US-09-328-111-529/c  
Sequence 529, Application US/09328111  
Patent No. 6262333  
GENERAL INFORMATION:  
APPLICANT: Endege, Wilson O.  
APPLICANT: Steinmann, Kathleen E.  
APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell III, Eddie  
APPLICANT: Carroll III, Theodore J.  
APPLICANT: Catino, Theodore J.  
APPLICANT: Derti, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: Monahan, John E.  
APPLICANT: Schlegel, Robert  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
FILE REFERENCE: CCD-257 (US)  
CURRENT FILING DATE: 1999-06-08  
CURRENT APPLICATION NUMBER: US/09/328,111  
EARLIER FILING DATE: 1998-06-10  
EARLIER APPLICATION NUMBER: US 60/088,801  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 529  
LENGTH: 251  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-328-111-529

Query Match 24.2% Score 32; DB 4; Length 251;  
Best Local Similarity 55.4% Pred. No. 0.11;  
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtctgacgagatttcgcgttgatctggtcttcacatcgagcttcgcg 64  
132 TAGAAGCCTGAGACATCTCTGCGCCGCAATCATCTAGTCTCATCTGCGCCCTCATCC 73  
DB 72 TAGCATTCTTACATFACACAGAGGATCAATCCCTCCCTTACCATCA 21

RESULT 2  
US-09-385-982-410  
Sequence 410, Application US/09385982  
Patent No. 6262334

```

GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: PRODUCTS: II
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-06
EARLIER APPLICATION NUMBER: 60/217,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 344
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 410
LENGTH: 353
TYPE: DNA
ORGANISM: Homo sapiens
US-09-385-982-410
```

```

Query Match          24.2%  Score 32; DB 4; Length 353;
Best Local Similarity 55.4%  Pred. No. 0.12;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgcttgagcagatgttcgcggtgtgtatctgtgtctcatcagtgcttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 tagaaaccgtctgacatctctccgcacatctctcctcctcctccatccc 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 65 tgcgcgtctgttcaaacagcagagttcccccagaggtgacatgcgtaccaa 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 tacgcatccttaccataaacagcagaggttcaacgcatcctccctaccacaa 330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 3
US-08-219-842-2
Sequence 2, Application US/08219842
Patent No. 5565323
GENERAL INFORMATION:
APPLICANT: Herrstadt, Corinna
TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
NUMBER OF INVENTIONS: for Alzheimer's Disease
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,842
FILING DATE: 30-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-AG 9504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 754 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
```

```

TOPOLOGY: linear
US-08-219-842-2
```

```

Query Match          24.2%  Score 32; DB 1; Length 754;
Best Local Similarity 55.4%  Pred. No. 0.16;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgcttgagcagatgttcgcggtgtgtatctgtgtctcatcagtgcttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 TAGAAACCGTCTGACATCTCTCCGCCCATCATCTAGTCTCATCTCCCATCC 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 65 tgcgcgtctgttcaaacagcagagttcccccagaggtgacatgcgtaccaa 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 TACGATCTTACATTAACAGAGGTCAACGATCTCTCCCTTACATCA 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 4
US-08-451-096-2
Sequence 2, Application US/08451096
Patent No. 5760205
GENERAL INFORMATION:
APPLICANT: Parker, W. D.
TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
NUMBER OF INVENTIONS: for Alzheimer's Disease
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,096
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/219,842
FILING DATE: 30-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-AG 9504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 754 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-451-096-2
```

```

Query Match          24.2%  Score 32; DB 1; Length 754;
Best Local Similarity 55.4%  Pred. No. 0.16;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgcttgagcagatgttcgcggtgtgtatctgtgtctcatcagtgcttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 TAGAAACCGTCTGACATCTCTCCGCCCATCATCTAGTCTCATCTCCCATCC 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 65 tgcgcgtctgttcaaacagcagagttcccccagaggtgacatgcgtaccaa 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 TACGATCTTACATTAACAGAGGTCAACGATCTCTCCCTTACATCA 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 5
US-09-385-982-539/c
Sequence 539, Application US/093855982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: EDEGEC, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-2609X
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 539
LENGTH: 789
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(789)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-539

```

TELEFAX: (202) 429-0796  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 854 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-04063-2

Query Match  
Best Local Similarity 24.2%; Score 32; DB 5; Length 854;  
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgtcgtgacgagttcgcgtgtgattcgtgtcattcgtgtcgtcgc 64  
DB 252 TAGAAGCGTCTGAACTCTCTGCGCCGATCATCTAGTCTCATGCGCCCTCCATCC 311  
OY 65 tgcgggtcgttcaaacagcagagttcccgagggtgacatcgtcatcaa 116  
DB 312 TACGCATCCTTTACATTAACAGACGAGTCAACGATCCCTTACCATCAA 363

## RESULT 8

US-09-097-889-2  
Sequence 2, Application US/09097889  
Patent No. 6218117

## GENERAL INFORMATION:

APPLICANT: Herinstdt, Corrina  
APPLICANT: Ghosh, Soumitra S.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING  
TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE  
NUMBER OF SEQUENCES: 26  
CURRENT APPLICATION DATA: EXTRA MITOCHONDRIAL DNA RATIOS  
FILING DATE: 15-JUN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman Ph.D., Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 660088.417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16569 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-097-889-2

ADDRESS: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 15-JUN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman Ph.D., Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 660088.417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16569 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-097-889-2

Query Match  
Best Local Similarity 24.2%; Score 32; DB 4; Length 16569;  
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgtcgtgacgagttcgcgtgtgattcgtgtcattcgtgtcgtcgc 64  
DB 7767 TAGAAGCGTCTGAACTCTCTGCGCCGATCATCTAGTCTCATGCGCCCTCCATCC 7826  
OY 65 tgcgggtcgttcaaacagcagagttcccgagggtgacatcgtcatcaa 116  
DB 7827 TACGCATCCTTTACATTAACAGACGAGTCAACGATCCCTTACCATCAA 7878

## RESULT 9

US-09-097-889-1  
Sequence 1, Application US/09097889  
Patent No. 6218117

## GENERAL INFORMATION:

APPLICANT: Herinstdt, Corrina  
APPLICANT: Ghosh, Soumitra S.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING  
TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE  
NUMBER OF SEQUENCES: 26  
CURRENT APPLICATION DATA: EXTRA MITOCHONDRIAL DNA RATIOS  
FILING DATE: 15-JUN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman Ph.D., Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 660088.417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6744 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-097-889-1

ADDRESS: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 15-JUN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman Ph.D., Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 660088.417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6744 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-097-889-1

Query Match  
Best Local Similarity 23.6%; Score 31.2; DB 4; Length 6744;  
Matches 60; Conservative 2; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgtcgtgacgagttcgcgtgtgattcgtgtcattcgtgtcgtcgc 64  
DB 4343 TAGAAGCGTCTGAACTCTCTGCGCCGATCATCTAGTCTCATGCGCCCTCCATCC 4402  
OY 65 tgcgggtcgttcaaacagcagagttcccgagggtgacatcgtcatcaa 116  
DB 4403 TACGCATCCTTTACATTAACAGACGAGTCAACGATCCCTTACCATCAA 4454

## RESULT 10

US-09-156-979-1  
Sequence 1, Application US/09156979  
Patent No. 5962672  
GENERAL INFORMATION:  
APPLICANT: Cowser, Lex M.

US-08-125-468-1

QY 62 cgctgccggtgctgttccaacagcaagagttcccgagggatgacatcgtcatcaac

2170 28881cc1gacgacgtgcttcgagcgccagctgctgtcctccttaagg 642

00-800-644-93

[illegible]



```

RESULT 15
US-08-776-210-3
: Sequence 3, Application US/08776210
: Patent No. 5942659
: GENERAL INFORMATION:
: APPLICANT: ALIBERT, Gilbert
: APPLICANT: MOULOUNGUI, Zephirin
: APPLICANT: BOUDET, Alain
: TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR
: DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: YOUNG & THOMPSON
: STREET: 745 South 23rd Street
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/776,210
: FILING DATE: 24-JAN-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 94 09272
: FILING DATE: 25-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/FR95/00957
: FILING DATE: 18-JUL-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: PATCH, Andrew J
: REGISTRATION NUMBER: 32,925
: REFERENCE/DOCKET NUMBER: IN 387 - BE 6996
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-521-2297
: TELEFAX: 703-685-0573
: TELEX: 248425 EMBON
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1600 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-08-776-210-3

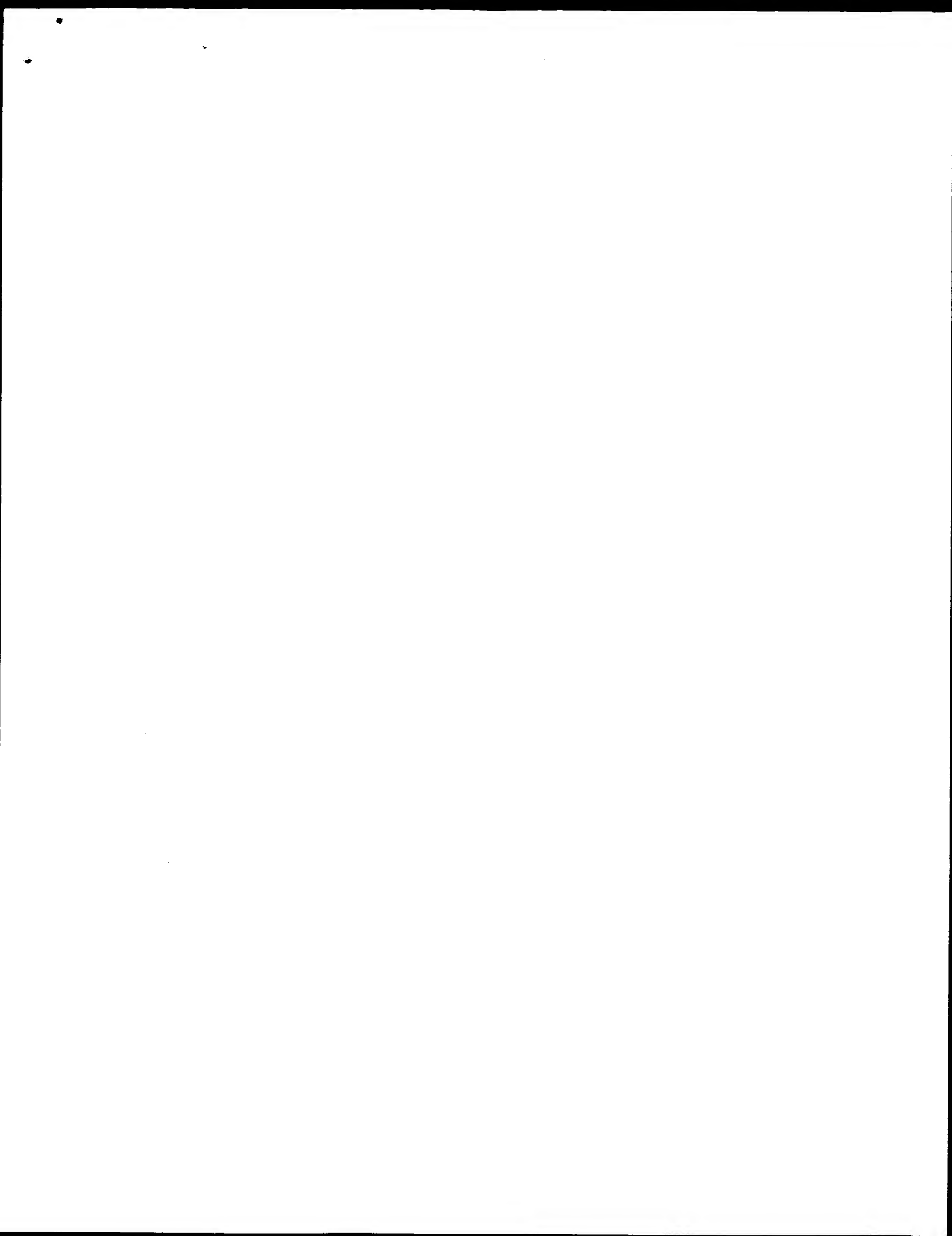
Query Match 20.9%; Score 27.6; DB 2; Length 1600;
Best Local Similarity 63.6%; Pred. No. 5.3;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 57 gtctcgctgcccgtgcttcaacaagaagttcccgaggtgagcagtcacatcaaa 116
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 894 gtgcgtcgctggcgtgcacgaacaaacccacagagtcacacacgcatatcgtagcgtt 953
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 117 cgtcga 122
   | |||
DB 954 CAACGA 959

```

Search completed: December 22, 2001, 09:36:27  
Job time: 5198 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 08:48:25 ; Search time 2235.47 Seconds  
(without alignments)  
634.517 Million cell updates/sec

Title: US-09-712-768-3  
Perfect score: 132  
Sequence: 1 ccgcgcgaatcgtctgac.....tcaacgtcgaaggtctgac 132

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estln: \*  
4: em\_estom: \*  
5: em\_estpl: \*  
6: em\_estba: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_hlc: \*  
10: gb\_est1: \*  
11: gb\_est2: \*  
12: gb\_hlc: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rtd: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	36.2	27.4	582	11	B1180715 LRRH22B7 L
2	36.2	27.4	643	11	B1180826 312287 LR
3	35.2	26.7	546	10	AM270568 x82807.x
4	35.2	26.7	640	10	AV702099 AV702099
5	34.4	26.1	244	11	W03206
6	34.4	26.1	283	10	AA176518
7	34.4	25.8	588	11	BG901209 950001C09
8	33.6	25.5	162	10	BE175243
9	33.6	25.5	177	10	BE180750
10	33.6	25.5	191	11	BG876169
11	33.6	25.5	198	11	BG876149
12	33.6	25.5	200	11	BG876153

Result No.	Score	Query Match	Length DB	ID	Description
13	33.6	25.5	202	11	BG876161
14	33.6	25.5	205	11	BG876168
15	33.6	25.5	258	11	BG875536
16	33.6	25.5	264	11	BG877167
17	33.6	25.5	295	11	C17077
18	33.6	25.5	303	10	AA879300
19	33.6	25.5	325	11	BF174777
20	33.6	25.5	361	11	BF170586
21	33.6	25.5	403	11	BG876871
22	33.6	25.5	462	11	BF172424
23	33.6	25.5	525	11	C17914
24	33.6	25.5	546	11	BG875704
25	33.6	25.5	554	10	A1028076
26	33.6	25.5	625	10	AV711246
27	33.6	25.5	662	10	AV727896
28	33.6	25.5	668	10	AV713349
29	33.6	25.5	688	10	AV762348
30	33.6	25.5	731	10	BG529384
31	33.6	25.3	536	11	BE971012
32	33.6	25.3	691	10	AV701188
33	33.6	25.3	716	10	A1525178
34	33.4	25.3	861	11	BE878356
35	33.2	25.2	187	10	AA092360
36	33.2	25.2	246	11	B1032317
37	33.2	25.2	339	10	AA179137
38	33.2	25.2	517	10	AA196841
39	33.2	25.2	776	11	BG116705
40	33.2	25.0	277	10	BE044951
41	33.2	25.0	738	11	BG612163
42	32.6	24.7	854	13	CNS04005
43	32.6	24.7	198	11	BG876186
44	32.6	24.7	323	10	A1160825
45	32.6	24.7	364	10	AA229353

#### ALIGNMENTS

RESULT 1  
LOCUS B1180715 582 bp mRNA EST 09-JUL-2001  
DEFINITION LRRH22B7 LRH (Lin Rhodamine and Hoechst dye) Mus musculus cDNA clone  
ACCESSION 3217089(5'), mRNA sequence.  
VERSION B1180715  
KEYWORDS B1180715.1 GI:14646526  
SOURCE EST.  
ORGANISM house mouse.  
REFERENCE Mus musculus  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 582) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
Ma, X.Y., Husain, T., Peng, H., Lin, S., Lu, B.F., Tuck, D., Mironenko, O.,  
Johnson, S., Krause, D. and Perkins, A.  
TITLE Development of a murine myeloid cDNA microarray using a subtracted  
JOURNAL CDNA library  
COMMENT Unpublished (2001)  
CONTACT: Hui Peng, MD, & Phd.  
Dr. Archibald Perkins' lab  
Yale University School of Medicine  
310 Cedar Street, LH 305, New Haven, CT 06520, USA  
Tel: 203-764-9977  
Email: hui\_peng@yale.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: M13A6K as 3' primer, 17 as 5' primer.  
Location/Qualifiers  
1. 582  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/clone="3217089(5)"  
/clone\_lib="LRH (Lin Rhodamine and Hoechst dye)"  
/sex="Female"





RESULT	6
AA176518	
LOCUS	
DEFINITION	AA176518 283 bp mRNA EST 09-MAR-1998
ACCESSION	AA176518
VERSION	AA176518.1 GI:1757651
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 283)
AUTHORS	Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kitzman,D., Kucaba,T., Lacey,M., Le,N., Lennon,G., Merita,M., Mertz, J., Moore,B., Schellenberg,R., Stepleso,M., Tan,F., Theisling,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE	Mashu-NCI human EST Project
JOURNAL	Unpublished (1997)
COMMENT	Contact: Wilson RK  National Institute School of Medicine

## FEATURES

## Source

```

/organism="Homo sapiens"
/db_xref="GDB:4643436"
/db_xref="taxon:9606"
/clone="IMAGE:611597"
/clone_lib="Stratagene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: bluescript SK-;
Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally.
Primer: Oligo dT. Skeletal muscle: from patient with
malignant hyperthermia. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGGAG
3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'"

```

Query Match	26.1%	Score 34.4	DB 10	Length 283
Best Local Similarity	55.4%	Pred. No. 4.8		
Matches	62	Conservative	0	Mismatches 50
				Indels 0
				Gaps

Qy	5	tyggaatcgtctgtagacatgtgtccggtgtgtautctgattcttcacatcgatcggtcgtctcgc	64
Db	84	TAGAAACCTCTGAACTATCTCCGCCGNCATCATCTCTATCTATCCGCCCTCCCATCC	143
Qy	65	tcccggtgtgtcttaacagcgaagatcccggaagggtgacatcgtcatcaa	116
Db	144	TACGCATCTTTNCATACACAGACGAGGTCAACGATCCCTCCTTMCATCAA	195

RESULT	7			
LOCUS	BG901209/c	588 bp	mRNA	EST
DEFINITION	95.0001C09. x2 950 - Mature pollen from Sheila McCormick's lab Zea			
ACCESSION	BG901209			
VERSION	BG901209.1	GI:14311454		
KEYWORDS	EST			

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
zea mays.	zea mays.	1 (bases 1 to 588)	Walbot,V.	Maize ESTs from various cDNA libraries sequenced at Stanford University	Unpublished (1999)	Contact: Walbot V
Eukarya: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC clade: Panicoideae: Andropogoneae: Zea.						Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 950001 row: C column: 09.
FEATURES						location/Qualifiers
SOURCE						1..588

## FEATURES

## Source

```

/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="950 - Mature pollen from Shelia McCormick's
lab"
/tissue_type="pollen"
/dev_stage="mature"
/lab_host="SOLR"
/notes="Vector: Stratagene's uni-zap XR (pBluescript SK-);
Site.1: EcoRI (5'-prime); Site.2: XhoI (3'-prime);
Unamplified cDNA library directionally cloned by Rima
Kulkasukas using Stratagene's uni-zap system. Insert
sizes ranged from 0.5kb to 2kb. 50 microliter aliquot had
338,000 pfu when it was made in sept, 1995, from oligo
dT-primed poly A+ RNA."
BASE COUNT      126 a      194 c      169 g      99 t
ORIGIN

```

ORIGIN

126 a	194 c	169 g	99 t
-------	-------	-------	------

Query Match	25.8%;	Score 34;	DB 11;	Length 588;
Best Local Similarity	63.4%;	Pred. No. 7.1;		
Matches 52;	Conservative 0;	Mismatches 30;	Indels 0;	Gaps 0;

QY	43	gtcttcacgtagcgtcttcgtctccaggtgctgtccaacgacgaagtgctcccgagagtc	102
Db	403	gccttcacatctctgcgcgcgcgacgttgacgagagtgctcaagaagatctccgcctcccgagagcc	344
QY	103	gacatcgtcatcaaatctgaag	124
Db	343	agcagcatgcgccacgctccggg	322

## RESULT

E175243

DEFINITION	Ov3-UTR0577-080500-311-c09 HT0577 Homo sapiens CDNA. mRNA sequence.
LOCUS	BEL/5243 162 bp mRNA EST Z1-JUN-2000

ACCESSION	BE175243
DEFINITION	QVZ-H10377 080300 211

VERSION BE175243.1 GI:8637969

**KEYWORDS** EST, human  
cortisol

ORGANISM	Homo sapiens
SOURCE	Human.

Eukaryota; Metazoa; Cr

Mammalia; Eutheria; Placentalia (basal to 1 to 163)

1 (bases 1 to 102)

Nagai, M. A., da Silva, W.

Goldman, G. H., Carvalho

BRUNSCHEIN, A.; SOARES, F.; BRUNSCHEIN, A.

Simpson, A.J.

TITLE Shotgun sequencing of

	sequence tags	Sci
TOTPNAT	proc Natl Acad	

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Prof. Antonio Prudente 109, 4 andar, 01509-010, Sa

This sequence was derived from the project. This entry can be seen

This sequence was derived from the project. This entry can be seen

200-101-e09&t3-2000-02-23&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 191.  
Location/Qualifiers

## FEATURES

source

1. 191  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0352"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 58 a 66 c 25 g 41 t 1 others  
ORIGIN

Query Match 25.5%; Score 33.6; DB 11; Length 191;  
Best Local Similarity 56.2%; Pred. No. 7.6;  
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 5 tggaaatcgtctgagcagatgttcgcgtgtgattcgtcttcacgagtgctcgc 64  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 41 TGGAAACCGTCTGACATACGAGAGTCCGCCCATCATCTTACTTATGCCCTCCATCCC 100  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
OY 65 tgcgcgtgtctgttcaaacagagagttcccgagagtgatcatcgtatcaaa 116  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 101 TACGCATCTTACATACAGAGAGTCCGCCCATCATCTTACTTATGCCCTCCATCCC 152  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

## RESULT 11

BG876149 198 bp mRNA EST 30-MAY-2001  
LOCUS BG876149  
DEFINITION MRI-CT0352-220200-102-a04 CT0352 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG876149  
VERSION BG876149.1 GI:14253239  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 198)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=6t2-MRI-CT0352-220  
200-102-a04&t3=2000-02-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 198.  
Location/Qualifiers

FEATURES  
source 1. 198

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0352"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 58 a 68 c 29 g 43 t  
ORIGIN

Query Match 25.5%; Score 33.6; DB 11; Length 198;  
Best Local Similarity 56.2%; Pred. No. 7.7;  
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 5 tggaaatcgtctgagcagatgttcgcgtgtgattcgtcttcacgagtgctcgc 64  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 48 TGGAAACCGTCTGACATACGAGAGTCCGCCCATCATCTTACTTATGCCCTCCATCCC 107  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
OY 65 tgcgcgtgtctgttcaaacagagagttcccgagagtgatcatcgtatcaaa 116  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 108 TACGCATCTTACATACAGAGAGTCCGCCCATCATCTTACTTATGCCCTCCATCCC 159  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

## RESULT 12

BG876153 200 bp mRNA EST 30-MAY-2001  
LOCUS BG876153  
DEFINITION MRI-CT0352-220200-102-c07 CT0352 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG876153  
VERSION BG876153.1 GI:14253243  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 200)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=6t2-MRI-CT0352-220  
200-102-c07&t3=2000-02-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 42  
High quality sequence stop: 200.  
Location/Qualifiers

FEATURES  
source 1. 200

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0352"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived



ORIGIN				
Query Match	25.5%;	Score 33.6;	DB 11;	Length 202;
Best local Similarity	56.2%;	Pred. No. 7.7;		
Matches	63;	Conservative	0;	Mismatches 49;
			Indels	0;
			Gaps	0;

49    5    tggaaatcgtcgcgcgacattgtccggtttgtatctcgtcctcaatcggatgcgtctcgc 64  
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 52    TGGAAACGCTTGACACTTCTCTGCCGCCATCATCTAGTCTCATCGCCCTCCATCCC 111  
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

112 TACGCATCCTTACATATACAGACGAGGTACACATCCCTTACACATCAA 163

RESULT 14			
BG876168/c			
LOCUS	BG876168	205 bp	MRNA
DEFINITION	MR1-CT0352-230200-101-d11 CT0352 Homo sapiens CDNA, mRNA sequence	EST	30-MAY-2001

RESULT	14
LOCUS	BG876168/c
DEFINITION	MG876168
ACCESSION	MG876168
VERSION	MG876168.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 205)
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R. Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1 (bases 1 to 205)	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A.	97 (7), 3491-3496 (2000)	
20202663	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil				

```

COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=ct2=WR1-CT0352-230)
200-101-d16t3=2000-02-23&ct=1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 205.

FEATURES
         source
             1..205
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone_id="CT0352"
                 /dev_stage="draft"

```

```

FEATURES
source
location/Qualifiers
1..205
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="CM0352"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:

```

Query Match	25.58;	Score 33.6;	DB 11;	Length 205;
Best Local Similarity	56.28;	Pred. No. 7.7;		
Matches	63;	Conservative	40	

Qy 5 tggaaatcgtctggaacgattgtccggtgtgtgattcgtgttcaatcgtgtcgtcgc 64  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 165 TGGAAACCGTCTGAACTATCCCTGCGCCATCATCTCTCATTCGCTCCATCC 106  
 Qy 65 tggcgggtcgtctgcaacgcaagagttcccggaagggtgaacatcgtcatca 116  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 105 TACGCATCCTTTACATACAGACGAGTCAACGATCCCTCCCTTACATCA 54

## RESULT 15

BG875536/c

LOCUS BG875536 258 bp mRNA EST 30-MAY-2001  
 DEFINITION RC6-CN0009-020200-011-F12 CN0009 Homo sapiens CDNA, mRNA sequence.

ACCESSION BG875536

VERSION BG875536.1 GI:14252520

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 258)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=st2-RC6-CN0009-020>)

200-011-F12st3-2000-02-02st4-1)

Seq primer: puc 18 forward

High quality sequence stop: 258.

Location/Qualifiers

1..258

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CN0009"

/dev\_stage="Adult"

/note="Organ: colon; normal: Vector: puc18; Site: 1: SmaI;  
 Site: 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

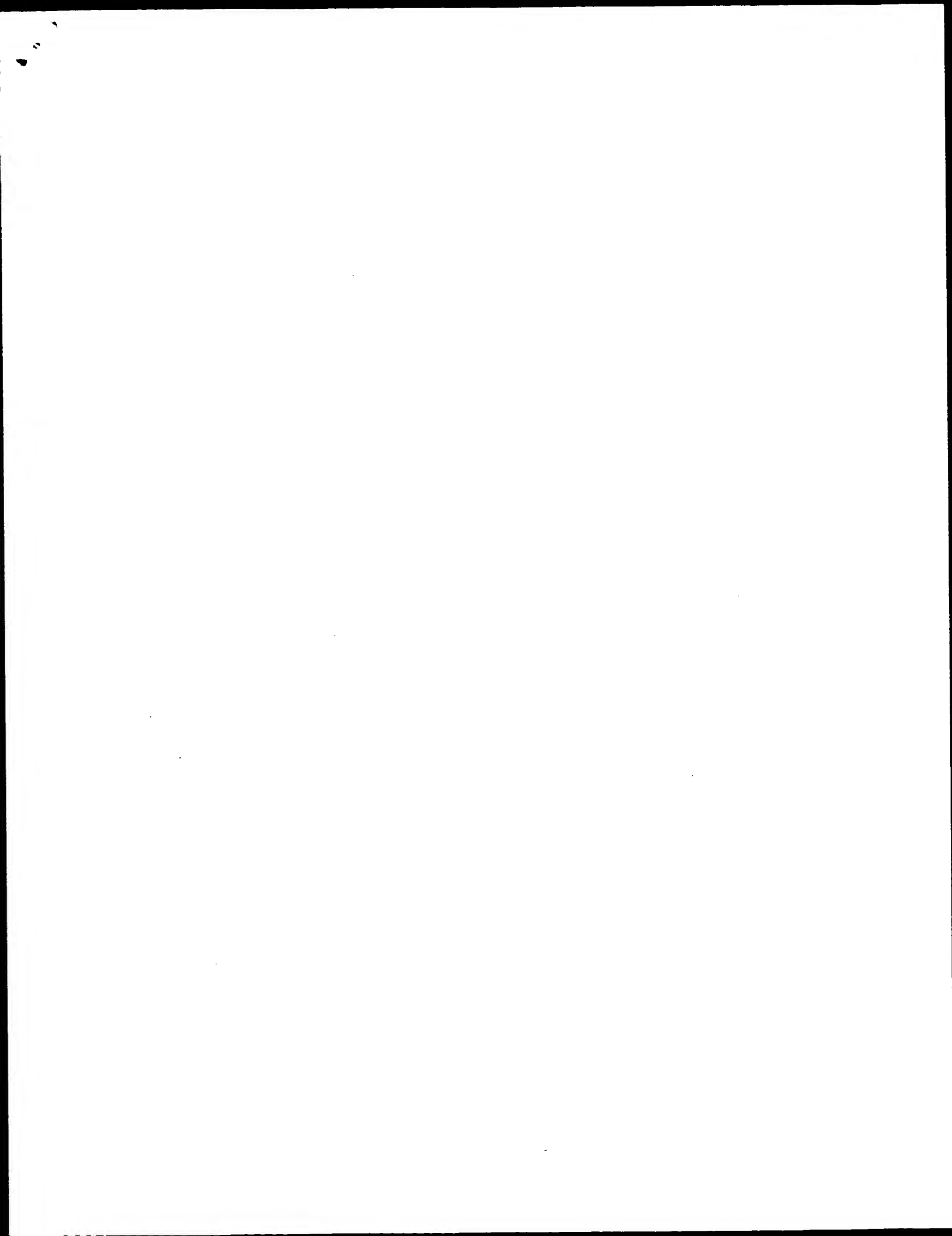
Db 152 TACGCATCCTTTACATACAGACGAGTCAACGATCCCTCCCTTACATCA 101

Search completed: December 22, 2001, 08:48:27  
 Job time: 2448 sec

Thu Dec 27 08:21:46 2001

us-09-712-768-3.rst

---





```

|||||
201 AGGCGACGTCGATCGCGGATGCGCTCGACAGACATGACGGCGAAGCGAC 250
84 1sleuTrpAsnValMetValThrTrpHisGlyIleLeuMetPhePhe 100
251 ACCTGTGGACGTCATGGTTACCTACCATGGTATTCGATGATGTTCTT 300
101 ValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTrpLeuMetProLe 117
301 GTGGGATACCCCGCATTTCTCGGTGTTTGGTAACTATCTGATGCCGCT 350
117 uGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeuSerP 134
351 GCAAAATCGGCGCTCGGATATGGCTTCCCGGTATGAAACACTGTCGT 400
134 hetPleuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAla 150
401 TCTGGCTGTCATTCGCGGTACCGCGATGGCGGTGGCTTCGCTTCCGCA 450
151 ProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTrpValLeuTrp 167
451 CCGGGGGGTGACGGTCACTGGGTTCCGGGCTTGGTTGGTTCTGTACCC 500
167 oProLeuSerThrArgGlnAlaGlyTrpSerMetAspLeuAlaIlePhe 184
501 GCCCTGTGACCGCGCAAGCTGGCTATTGATGGACCTCGCATTTTCG 550
184 1aValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIle 200
551 CGGTATCACTGTGGGTGCTCTCTGATCATAGGCGCGCATCAACATATC 600
201 ThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisValPr 217
601 ACGACCTTCTTGAACATGCGCGCCCGGCATGACGGTGCAAAAGTCC 650
217 oLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLeuAla 234
651 GTTGTTCGCGGTGATCTTATCACGGCTTGGCTATCTCTGTCGCGC 700
234 euProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPhe 250
701 TGCCGCTTCTGGCTGGTGCATACCATGCTGCTGACCGACCGCAACTTC 750
251 GlyThrThrPhePheAsnProAlaGlyGlyGlyAspProIleLeuTrp 267
751 GGCACGACCTTCTTCAATCTGCTGGCGGGGTGACCCGATTTGTACCA 800
267 nhIstIleLeuTrpPhePheGlyHisProGluValTrpIleIleLeu 284
801 ACACATCTCTGTGTTCTTGGGACCGCGAAGTGTACATCATCTTCTGC 850
284 roGlyPheGlyIleIleSerHisValAlaSerThrPheSerLysPro 300
851 CCGGCTTTGGCATCATGACCATGCTGTCGACCTTCTCGAAAAAACCG 900
301 ValPheGlyTrpLeuProMetValAlaAlaMetValAlaIleGlyVal 317
901 GTCTTCGGTACCTGCGGATGCTATGCAATGCTGCAATCGGTCTTCT 950
317 uGlyPheValValTrpAlaHisHisMetTrpTrpValGlyMetSerLeu 334
951 GGGCTTTGTCTGTGGCGCACACAGTACACCGTTGTGATGTCGTGA 1000
334 hngInGlnSerTrpPheMetLeuAlaThrMetValIleAlaValProThr 350
1001 CCCAGCATCTCTACTTCATGCTGGCCACCATGTCGATCGCGGTCCAGC 1050
351 GlyIleLeuHisIlePheSerTrpIleAlaThrMetTrpGlySerVal 367
1051 GGCATTTAAGATCTTCTCGTGATGCGCACATGTCGGCGGCTCGTTGA 1100
367 uPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPheTrp 384
|||||

```

```

1101 GTTCAATCGCGCATGCTGTGGGCTTTTGGCTTATGTTCTGTTACCG 1150
384 a1GlyGlyValIThrGlyIleValLeuAlaGlnAlaGlyLeuAspArgAla 400
1151 TGGGTGTTGTACCGGATGTCGTGCTGCGCCAAAGCGGCTCGACCTGCA 1200
401 TyrHisAspThrTrpTrpValValAlaHisPheHisTrpValMetSer 417
1201 TATCAGACACCTTATTACGTGGTGGCGCACTTCCATTATGTCGTGCT 1250
417 uGlyAlaIlePheAlaIlePheAlaGlyIleTrpPheTrpMetProLys 434
1251 GGGTGGATGCTTGGCATGCTTCCCGGTATCTACTTTTACATGCCGAAT 1300
434 heSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTrpThr 450
1301 TCTGGGCGCGGCTTCCCGGAATGGCGTGCAGAAAGCTGCACCTTCCGAC 1350
451 PhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuGly 467
1351 TTTCTCATGCTGTCGAAAGCTACAGCTTCTTCCGACACCTTCTGGAGC 1400
467 gGlnGlyMetProArgArgTrpIleAspTrpProGluAlaPheAlaLeu 484
1401 TCAGGGATGCGCGCGCGTTACATGACATGCCGAAGCTTCCGCGCTGT 1450
484 rPasLysValSerSerTrpGlyAlaPheLeuAlaPheAlaSerPheLeu 500
1451 GGAACAAGTCTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
501 PhePheIleValIlePheValTrpThrLeuValAlaGlyArgArgGly 517
1501 TTTCTCATGCTGATCTTGTGTATACGCTGCTGCTGCTGCTGCTGCTG 1550
517 rArgProAsnProTrpGlyGluPheAlaAspThrLeuGluTrpThrLeu 534
1551 CCGTCCGAACCGGTGGCGCAATTCGCCGATAGCGTGAATGACGACGTC 1600
534 roSerProProProAlaHisThrPheGluThrLeuProLysArgSerAsp 550
1601 CATCACCGGCTCGGCGCCACACCTTGGAACGCTGCCCAAGGCTGGAGC 1650
551 TrpAspLysHisProSerHis 557
1651 TGGGACAGAGATCCCTCGCAC 1671

seq_name: gb_ba:PCYCBA
seq_documentation_block:
LOCUS PCYCBA 2629 bp DNA BCT 23-JAN-1992
DEFINITION P.denitrificans cyc A and cta dII genes for cytochrome c 550 and
cytochrome c oxidase beta subunit I.
ACCESSION Y07533.1 GI:45477
VERSION Y07533
KEYWORDS cta dII gene; cycA gene; cytochrome c oxidase; cytochrome c-550.
SOURCE Paracoccus denitrificans.
ORGANISM Bacteria; proteobacteria; alpha subdivision; Rhodobacter group;
Paracoccus.
REFERENCE 1 (bases 1 to 2629)
AUTHORS Raitio,M.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1990) M. Raitio, DEPT OF MEDICAL CHEMISTRY,
UNIVERSITY OF HELSINKI, SILTAVUORENRENNER 10, 00170 HELSINKI 17,
FINLAND
2 (bases 1 to 2629)
AUTHORS Raitio,M., Pilsa,J.M., Metso,T. and Saraste,M.
TITLE Are there isoenzymes of cytochrome c oxidase in Paracoccus
denitrificans?
JOURNAL FEBS Lett. 261 (2), 431-435 (1990)
MEDLINE 90184495
FEATURES Location/Qualifiers
source 1..2629

```

```

/organism="Paracoccus denitrificans"
/scraim="PD1222"
/sub_strain="ssp. MR-3"
/db_xref="taxon:266"
/cione_id="EMBL3 lambda"
233.. 2544
/gene="cyc A"
233.. 292
233.. 700
/gene="cyc A"
233.. 700
/gene="cyc A"
/EC_number="1.9.3.1"
/note="cytochrome-c oxidase"
/codon_start=1
/transl_table=1
/product="cytochrome c 550"
/protein_id="CA68820.1"
/db_xref="GI:45478"
/db_xref="SWISS-PROT:P00096"
/translation="MKSIYATLALPAAQDDDAKGEFKNKCRACHIQAP
DGDITKGRTGPNLVGVGRKLASEGFYGBEILVAENPDLTEADILEVTD
PKPLVMKTDCKGAKTKMFKMRKNQADVAFILAQNPSDPAGDEAAEGSN"
293.. 2544
/gene="cyc A"
/EC_number="1.9.3.1"
/note="cytochrome c 550"
/product="cytochrome-c oxidase"
731.. 765
/gene="cyc A"
871.. 2547
/gene="cta DII"
871.. 2547
/gene="cta DII"
/EC_number="1.9.3.1"
/note="cytochrome-c oxidase"
/codon_start=1
/transl_table=1
/product="cytochrome c oxidase beta subunit I"
/protein_id="CA68821.1"
/db_xref="GI:45479"
/db_xref="SWISS-PROT:P98002"
/translation="MDAAVHGHDHDTRGFTRMPESTNHNKDIGLLPYTAGIVG
ISVEFYVMRELOHPGVGYMCIFLGARILADAECIPNGHLMMVTIYHGVLMEFE
VIRLPFGPGNYEMRLHIGARDMAFPRLNNISYMYVCVALGYASILLAPGNOMG
GVGAVLYPLSTEAGYSMDLAIFYHVSGASSILGAINITIITPLMKRPFMTLEKHA
LEFAWVFITAWLILLSLEVLGAITMLMDNFCTQEPDPAQGDPLYQHILMEFG
PEVYIILPGFGIISHVISTFAKPPIFGYLPMVLAMAIGILGEVYAHNMITYGMSI
TOOAYFMALTIAVPGIKVESFIATMGGSIEPKRPLMAFGFLFETGVGYGVYGL
LSQAPLDRVYHDHYVVYVAHFHYSLGAVESIFPGVYYXVIIGKMSGROYDEAQLHE
MFQGSNIIFPPQHLRGCPRIKIDYPVFAIWNINSIGATISPSFLFCIYIVA
YTTFAGKRVNPYNWEHADYLENTLSPPEHTPETLRKEDWDRAHAH"
2586.. 2615

terminator      457 a      872 c      771 g      529 t
BASE COUNT
ORIGIN

alignment_scores:
    Quality: 2538.50          Length: 558
    Ratio:   4.854            Gaps:     1
Percent Similarity: 93.728    Percent Identity: 81.004

alignment_block:
US-09-712-768-2 x PDCYCA ..
Align seg 1/1 to: PDCYCA from: 1 to: 2629

1 MetAlAsPALaLaLehniSgLyHiS...AsPhisHisgLuLySgingL 16
|||||.....:|||||||.....:|
871 ATGGCAGACCCAGACCCTTCACGCCACGGTGACCATCATCAACACCCCGG 920
yphePhetHarTirPhemetSerThranhiSlYsAsPlLgLyLeuL 33
|||||.....:|||||||.....:|
921 GTTCTCACCCCGGTGTTCTATGTCAACAACCAACAGATATCGGTATCC 970

```

33	euYrLeuValAlaAlaValValValAlaGlyPheIleSerValLeuPheThr	49
971	TTTTACTGTTTACAGGGCGGACATCGTGGCCGATCTCGATTGCTTACC	1020
50	ValTYrMetArgLeuGluLeuMetAspProGlyValGlnTYrMetCysLe	66
1021	GTCATTATGCGCATCGAAGCACTGGACATCCGGGCGCTGCAGATACATGTGCTT	1070
66	uGluGlyAlaArgLeuIleAlaAspAlaSerGlnThrCysThrLAsnG	83
1071	GGAGAGCGCGCTTCATTCGCGAGCGCTCGGCGAATGACCCCGAGAC	1120
83	YHisLeuTrpAsnAlaMetValThrTrnHisGlyIleLeuMetCLeuPhe	99
1121	GACACCTGGGAGAACTCATGATACACTTACCAAGCGGTGCTCATGTATTC	1170
100	PheValGlyIleProAlaLeuPheGlyIlePheGlyAsnTYrLeuMetLr	116
1171	TTTCGTGTATCCCGGCACTGTTCGGCGGTTTCGGCAACTTTTTCATGCC	1220
116	oleuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeuS	133
1221	GCTGCATTCAGCGCGCCCGGACATGGCTTCCCGCGGTGAACAACCTCT	1270
133	erPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPhe	149
1271	CCTACGATGATATGTCATCGCGCGCTGGCGCTGGCGTCCGCTGCTG	1320
150	AlaProGlyGlyAspGlyIleuLeuSerGlyValaGlyTYrPValLeuTY	166
1321	GGCGCGGGGGGCAACCAAGATGGGTTTGCGGCTGGCTGGGTGCTCTA	1370
166	rProProLeuSerThrArgGluAlaGlyTYrSerMetAspLeuAlaIleP	183
1371	CCGCGCGCTTCGACACCAAGCGGCGGCTATTCATGCATGGCGATCT	1420
183	heAlaValAlaHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMet	199
1421	TTTCGCGTCCACGCTCTGGGTGCTCTGTCGATCTGGGCGCATCAACATTC	1470
200	IleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisLysVala	216
1471	ATACACCACTTCCTCTCAACATGGCGGACCGGCGCATAGACGTGTTCAAGT	1520
216	IProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLeuLeuA	233
1521	GGCGCTGTGTGCTGGCTGGCTTCATCAACCGCTCGGTGATCTCGTGT	1570
233	IleuSerProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsn	249
1571	CGGTGCGGTTCTGGCGGCGGCGGATCCACATCTCTCATGGACCGCAAC	1620
250	PheGlyTYrThrPhePheAsnProAlaGlyGlyAspProIleLeuTY	266
1621	TTGCGACAGAGTTTCTTCATCCGCGGCGGCGGCGACCGGTCCTTTA	1670
266	rgIlnHisIleLeuTrpPhePheGlyHisProValTYrIleIleIleIle	283
1671	CCAGCAACATCCGTGGTCTTTCGGCATCCGAGGTCTATATCAATCAATCC	1720
283	eufProGlyPheGlyIleIleSerHisValaValSerThrPheSerLysLys	299
1721	TGGCGGGCTTGGSCATCATCAGCAACGATCATCTGCACCTTGGCCANAAG	1770
300	ProValPheGlyTYrLeuProMetValTYrAlaMetValAlaIleGlyVala	316
1771	CCGATCTTGGCTACCTGCGCATGTGGTGGCGGCATAGGCGGATCGCAT	1820
316	IleuGlyIlePheValValTrpAlaHisHisMetTYrThrValaGlyMetSerL	333
1821	CTGGGGTCTGTGCTGTGGGCGGACCAATGTACACGGCGGCAATGTGCG	1870

333	euThrGlnInseryrPheMetLeuAlaThrMetValLLaValaPro	349
1871	TeACCCAGAGGCGCATTTCATGCTGGCGACCATACCATGCGGTGCC	1920
350	ThrGlyILeLysILePheSerTrpIleAlaThrMetTrpGlyGlySerVa	366
1921	ACCGGATCAAGGATCTTCCTCGATATCGGACCATGATGGGCGGAGCAT	1970
366	LcIuPheLysSerTrpMetLeuThrAlaPheGlyIuPheMetPheLeuPheT	383
1971	CGAGTTCAGAGCGGAGATGCTGGCGCTTGGCTTCTGTCTCTGTTCCATCA	2020
383	hVaILeGlyValThrGlyILeValLeuAlaGlnAlaGlyLeuAspArg	399
2021	CGGTGGGGGGGTGACCGCGCGTGGTGCAGACGACGCGCGCGGTGACCGG	2070
400	AlaTyRHisAspTrhTyTyrValValAlaHisPheHisTyRValMetSe	416
2071	GTCTATCAACAGACCATTAATACCTCGTGGCCCACTTCCACTAGCATGTCT	2120
416	IleuGlyAlaIlePheAlaIlePheAlaGlyILeIuPheIuPheIuMetProL	433
2121	GCTGGGCGCGTGTTCGGGCAATCTTGGCGGGGGTCTATTACGTGGATCGGCA	2170
433	ysPheSerGlyArgAlaIlePheProGluTrpAlaAlaLysLeuHisPheTrp	449
2171	AGATTCGGGCGCGGCAATACCGGAAATGGCGGGGCGACATGATTTGTGG	2220
450	ThrPhePheILeGlyAlaAsnValThrPhePheProGlnHisPheLeuG	466
2221	ATGATGTTTCATCGGTCTGCAACCTGTATCTTCTTCCCGGACGACTTCCGTGG	2270
466	YArgGlnILeMetProLArgArgTyrILeAspTyrProGluIuPheAlaL	483
2271	CGCGCAGGCGCATGGCGCGCGCTAATGCACTATTCGGTGCAGTTCCGCT	2320
483	euTrpAsnLysValSerSerTyRGLyAlaPheLeuAlaPheAlaSerPhe	499
2321	ATTGGAAACAATCTCGTCGATTCGGCGCGCTAATCTCTTCCGCTCTCTTC	2370
500	LeuPhePheILeValILePheValTyRThrIleuValAlaGlyArgArgGL	516
2371	CTGTTCTTCATATGGCATCGTCTTACACGCTCTTCCGCGGCAAGCGCGT	2420
516	uThrArgProAsnProTrpGlyGluPheAlaAspTrhLeuGluIuTrpThrL	533
2421	GAACTGCGCAACTAAGTGAACGAGCATGCGGACACCTGTGAATGAGACC	2470
533	euProSerProProAlaHisThrPheGluIuTrhLeuProLysArgSer	549
2471	TGCCCTCCCGCCGCGGAGCATACCTTCGAGACGACCTGCGCAAGCGCGAG	2520
550	AspTrpAspLysHisProSerHis	557
3521	GACTGGGATCGCGACACGCGCAT	2544

seq_name:	gb_ba:RSCOXS1
seq_documentation_block:	
LOCUS	RSCOXS1 2166 bp DNA
DEFINITION	R. sphaeroides gene for cytochrome oxidase subunit 1.
ACCESSION	X62645
VERSION	X62645.1 GI:46417
KEYWORDS	cytochrome oxidase subunit 1.
SOURCE	Rhodobacter sphaeroides.
ORGANISM	Rhodobacter sphaeroides. Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group Rhodobacter.
REFERENCE	1 (bases 1 to 2166)
AUTHORS	Shapleigh,J.P.
TITLE	Direct Submission
JOURNAL	Submitted (17-OCT-1991) J.P. Shapleigh, University of Illinois Dept of Biochemistry, 505 S Mathews, Urbana IL 61801, USA

2 (bases 1 to 2166)  
 Shapleigh J.P. and Gemm R.B.  
 Cloning, sequencing and deletion from the chromosome of the gene  
 encoding subunit I of the aa3-type cytochrome c oxidase of  
 Rhodobacter sphaeroides  
 Mol. Microbiol. 6 (5), 635-642 (1992)  
 92204019

JOURNAL  
 MEDLINE  
 FEATURES  
 SOURCE

CDS

326 a 746 c 639 g 455 t

Location/Qualifiers  
 1..2166  
 /organism="Rhodobacter sphaeroides"  
 /strain="GA"  
 /db\_xref="taxon:1063"  
 207..1904  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome oxidase subunit 1"  
 /protein\_id="CAA4514.1"  
 /db\_xref="GI:46418"  
 /db\_xref="SWISS-PROT:P33517"  
 /translation="MADAAIGHGHEHRRGFFTEFMSTNHHKIDIGVLYFTGTGLILGILISL  
 VAFVYVMELMARPQVOMCAENHESGLVAFSPQSLMPCGSAVEICNPCHNCHVMVLTGSG  
 GLIMEFVYVIALFGGNGNYEMPHILGAPMAAPRRNNISLYLYAAGTSLAVASLPAH  
 GNGNGLSGGIGWLYPRPLSTSESGTLDLITMVLHLSGASITLGAINMTYFLNNRAA  
 GMTNRPVLFAMSLFTVATWLLILALPVLAATITALLTDKNFTGFQSGGGDPLVLLK  
 HILMEFGHEVYIIVLPAGFIVSHVITFAKKPIFGYLDPMVYAAVIGLGVVANH  
 MTAALSTLQOYSFMMATNVAFTGKIKTSMTATMGSSIELKTPMLALGFLLEFLFGL  
 VAGVGIYLSQASVDRYHDYLYVVAHFHVNMSLAVFGFISFGSTSGIGKMSGRQYPR  
 VGGKTHPEMMEVGNILTFRRDHLFGROGMRRRIIDVPEAFATWNVSSSLGAPLSPAS  
 LEFLGVITYSLSGARVATANNVYMHADTLEWTLTSPRPHTFEQLPKRDMERAPAH"

alignment_scores:	Quality: 2453.00	length: 567
Ratio: 4.736	Gaps: 3	
Percent Similarity: 91.358	Percent Identity: 78.660	
alignment_block:		
US-09-712-768-2 x RSCOXSI ..		
Align seg 1/1	to: RSCOXSI	from: 1 to: 2166
1	MetaLaSPaLaLaLaLeNiSgLYNiASpRHiSgLUySgNgLYpH	17
207	ATGCGCGAGCGAGCATCATGCGCCAGAG...CACAGCCGAGGGGTT	253
17	ePhetHargTrpHemetSerThraSnHsLYsAspRLeGlyLeuLeut	34
254	CTTCACCCCGCTGTTCAATGTCAGAACCCAAAGAGCATCGGGCTTCT	303
34	YrLeuValAlAlaIGLYaValaIGlyheLleSeValLeuPhetHrVal	50
304	ATGCTTCACCGGGGGGCTCTCGGGCTGATCTCGGGCTTCACCGTC	353
51	TyrMeTarGLEuGIGuLeuMeTarSPrgLYaIGlyHrMeTysLeuGI	67
354	TACATCGCGAGTGGAGCTCATATGGCCGCCGGCGTGCATGTGGCCCG	403
67	uGlyLaArg.....LeuLeaLa	74
404	ACATGTGAGATCGGGCGTCGTAAGGGCTTCTTCAGATCCCTGTGGCCCT	453
74	SPaLaSeRGILThCysThrAlaAsnGLYHISLeuTrpAsnValMetVal	90
454	CGGCGTGGAGAAATCTGACCCCGAAGCCCATCTGTGGAAAGTCATGATGC	503
91	ThrTrpHISgLYIleLeuMetMetPheValaIGLYIeProLaLeuPh	107
504	ACCGGACCCGAGATCTGATGATCTTCTGTGTCATATCCCGCTCTT	553
107	eGlyGLYpHeGlysnLYIleuMetProLeuGInIleGlyAlaProAspM	124



554 CGCGGCTTCGGCAACTATTTTCATCCCGCTGCACATCGCGCGCGGACA 603  
 124 evlAlapheProArMethetAsnAnleuSerPheTrpLeuPheIleAlaGly 140  
 TGGCTTCCTCCGGATGAAACACCTCTCTACTGGCTCATGTCCGGGG 653  
 604 TGGCTTCCTCCGGATGAAACACCTCTCTACTGGCTCATGTCCGGGG 653  
 141 ThrAlaMetGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLe 157  
 157 uGlySerGlyValGlyTrpValLeuTrpProPheSerThrArgGlyAla 174  
 704 GGGCTTCGGCATCGCTGGCTGCTCATCCGCCCTCTCCACCTCGCAAT 753  
 174 laGlyTrSerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAla 190  
 754 CGGCTTCCTCCAGCATCGCATCTCGCGGTGCACCTGCGGGCGGCC 803  
 191 SerSerIleMetGlyAlaIleAsnMetIleThrTrpPheLeuAsnMetAr 207  
 804 TCCTGCATCCCTCGCGCATCAACATGATACGACCTCTCTGAACATGGC 853  
 207 gaIaProGlyMetThrLeuHisIleValProLeuPheSerTrpSerIleP 224  
 854 CGCACCGGCGCATGACATGCACAAAGTGGCGCTCTCGCTGTCGATCT 903  
 224 heIleThrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAlaGlyAla 240  
 904 TCGTACACCGCTGGCTGATCTGCTGGCGCTGCGCCGCTGCGCGCGCC 953  
 241 IleThrMetLeuLeuThrAspArgAsnPheGlyTrpThrPhePheAsnPr 257  
 954 ATTAGCATGCTGCTGACCGACCGACCTTGGCCACCACTTCTTCCAGCC 1003  
 257 oAlaGlyGlyAspProIleLeuTrpGlnHisIleLeuTrpPhePheG 274  
 1004 CTGCGGGGCGCGCGCGCTGCTCATACAGCATCTGCTGGTTCTTCG 1053  
 274 lYHisProGlyValTrpIleIleIleLeuProGlyPheGlyIleIleSer 290  
 1054 GCCACCGGAGGTCTACATCATCGCTGCGCGCTTGGCATGTGACG 1103  
 291 HisValSerThrPheSerIleValProValPheGlyTrpLeuProme 307  
 1104 CACGTCATGCGACCTTCCCAAAAGCCGATCTTGGCTTTTGGCGAT 1153  
 307 lValTrpAlaMetValAlaIleGlyValLeuGlyPheValValTrpAlaH 324  
 1154 GGTCTATGCGATGGTGGCATCGCGCTGCTGCTGCTGCTGCGGCC 1203  
 324 lHisMetTrpThrValGlyMetSerLeuThrGlnGlnSerTrpPheMet 340  
 1204 ACCACATGTATACACCGCGCTGACCTTACCCACAGACTACTTATATG 1253  
 341 leuAlaThrMetValIleAlaValProThrGlyIleIlePheSerTr 357  
 1254 ATGGCAGCATGTGTATCGACATGCCACCGCATCAAGATCTTCTCTG 1303  
 357 pIleAlaThrMetTrpGlyGlySerValGlnPheLeuSerPromeIleu 374  
 1304 GATCGGACCATGTGGGCGCTGATCGATCGACCAAGACGGGATGCTCT 1353  
 374 rpAlaPheGlyPheMetPheLeuPheThrValGlyGlyValTrpGlyIle 390  
 1354 GGGCGCTCGGCTCTCTCTCTCTTACCGTGGGCGGCTCACCGCATC 1403  
 391 ValLeuAlaGlnAlaGlyLeuAspArgAlaTrpHisAspThrTrpVal 407  
 1404 GTGCTGAGCGAGCGAGCGCTGACCGCTATTTATCAGACACCTACTATGT 1453  
 407 lValAlaHisPheHisTrpValMetSerLeuGlyAlaIlePheAlaIleP 424  
 1454 CGTGGCGCACTTCATTTATGTATGATGAGCTCGCGCGGCTTTCGCGATCT 1503

424 heAlaGlyIleTrpThrMetProIlePheSerGlyArgAlaPhePro 440  
 1504 TCCGAGATCTTACTTGTGGATCGGCACATCTCGCGCGCATATCCG 1553  
 441 GluTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGlyAlaAsnVa 457  
 1554 GAATGGCGCGGAAGCTCATTTCTGATGATGATTTCTGGGCGCAACT 1603  
 457 lThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProArgArgT 474  
 1604 CACCTTCTTCCGACACTTCTCGCGCCGACCAAGGATATGCGCGGCTT 1653  
 474 YrIleAspTrpProGlyAlaPheAlaLeuTrpAsnLysValSerSerTrp 490  
 1654 ACATGCACTTCCGAGGCTTGGCCACCTGGAACTTGTCTGTCGCGGT 1703  
 491 GlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleValIlePheVa 507  
 1704 GGGCTTCCTGCTCTTCCGCTGCTGCTCTCTCTCTCTCTCTCTCTCT 1753  
 507 lTrpThrLeuValAlaGlyArgGlyTrpThrArgProAsnProTrpGlyG 524  
 1754 CTACAGCTCG...AGCGGGCGCGGTGAGAGGAGCAACACTACTGAGACG 1800  
 524 lPheAlaAspThrLeuGlyTrpThrLeuProSerProProAlaHis 540  
 1801 AACATGCCGACACCTGAGTGGACGTGACCTCTCCGCGCGGAGCAT 1850  
 541 ThrPheGluThrLeuProIleArgSerAspTrpAspLysHisProserH 557  
 1851 ACGTTCGAGCACTTCCCAAGCGGAAGACTGGAGACCGCGCGCCCA 1900  
 557 s 557  
 1901 C 1901  
 seq\_name: gb\_da:PDCCOX2  
 seq\_documentation\_block: 3211 bp DNA BCT 07-OCT-1996  
 LOCUS PDCCOX2  
 DEFINITION P. denitrificans COI gene.  
 ACCESSION X05829  
 VERSION X05829.1 GI:994806  
 COI gene; cytochrome c oxidase; cytochrome oxidase a3;  
 KEYWORDS unidentified reading frame.  
 SOURCE  
 ORGANISM  
 Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 Paracoccus.  
 1 (bases 1 to 3211)  
 Raitio,M., Jalli,T. and Saraste,M.  
 Isolation and of the genes for cytochrome c oxidase in Paracoccus  
 denitrificans  
 EMBO J. 6, 2825-2833 (1987)  
 2 (bases 1 to 3211)  
 Saraste,M.  
 Direct Submission  
 Submitted (04-JAN-1991) Saraste M., EMBL, Meyerhofstrasse 1,  
 Heidelberg, Germany  
 revised by [3] MAT  
 3 (bases 1 to 3211)  
 Saraste,M.  
 Direct Submission  
 Submitted (20-SEP-1995) Saraste M., EMBL, Meyerhofstrasse 1,  
 Heidelberg, Germany  
 On Sep 21, 1995 this sequence version replaced gi:45474.  
 COMMENT  
 FEATURES  
 source  
 1. 3211  
 /organism="Paracoccus denitrificans"  
 /db\_xref="taxon:266"  
 /map="cox2 locus"  
 218. 222  
 RBS

[illegible]

```

363 GLYGISerValGluPheLysSerProMetLeuTrpAlaPheGlyPhe 379
1857 GGGGGCTCGATTCAGATTCAGACGCCGATGCTCTGGCCCTGGCC... 1901
379 tPheLeuPheTrpValGlyValThrGlyValLeuAlaGlnAlaG 396
1902 .TTCCGTTCACCGTCGGCGCGGTGACCGGGGTCGATCGCCACGAGG 1950
396 LysLeuAspArgAlaLysTrpHisAspThrTyrTyrValAlaHisPheHis 412
1951 CCTTGACCCCGTTTATACAGACACCTACTATATCGCGGCGATTTCCAC 2000
413 TyrValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyLeuTrp 429
2001 TATGTGATGACCTTGGCCCTCTTCGCGATCTTTCAGAGCACCTATTA 2050
429 eTyrMetProLysPheSerGlyArgAlaPheProGluTrpAlaAlaLysL 446
2051 CTGGATCGGACAGATGTCGGCGGCAATACCGGATGAGGCGGCGCAGC 2100
446 eHisPheTrpTrpPhePheIleGlyAlaAsnValThrPhePheProGln 462
2101 TGCATTTCTGATGATGATTCATCGGCTCGAACCTGATCTTCTCCGAG 2150
463 HisPheLeuGlyArgGlnGlyMetProArgArgTyrIleAspTyrProG 479
2151 CACTTCTGGGCGCGGCGGCGATGCGCGCGCTATATTCACACCGCGT 2200
479 uAlaPheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaP 496
2201 CGATCTCTCGACTGGAACAACATCTCGATCGGCGGCTATATCTCTCT 2250
496 heAlaSerPheLeuPhePheIleValIlePheValTyrTrpLeuValAla 512
2251 TCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2300
513 GLyArgArgGluTrpArgProAsnProTrpGlyGluPheAlaAspTrpLe 529
2301 GGCAGAGCCCGGAGACGTCGCGAACTACTGGAGACGATCGGACAGCGT 2350
529 uGluTrpThrLeuProSerProProPheAlaHisThrPheGluTrpLeu 546
2351 GGAATGACCTCGCTCCCTCGCGCGCGCGGACATACCTTCGAGACCTGCG 2400
546 rOLysArgSerAspTrpAspLysHisProSerHis 557
2401 CCAAGCCCTGAGGACTGGGATCGCGCACAGGCGCAT 2435

seq_name: gb_da:SME591785

seq_documentation_block:
LOCUS SME591785 286550 bp DNA BCT 16-AUG-2001
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 4/12.
ACCESSION AL591785 AL591688
VERSION AL591785.1 GI:15073719
KEYWORDS
SOURCE
ORGANISM Sinorhizobium meliloti.
Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
Rhizobiaceae: Sinorhizobium.
1 (bases 1 to 286550)
REFERENCE
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Bolstad,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
From the Cover: Analysis of the chromosome sequence of the legume
symbiont Sinorhizobium meliloti strain 1021
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9877-9882 (2001)
PUBMED 11481430
REFERENCE
AUTHORS Gouzy,J.
2 (bases 1 to 286550)

```

```

TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
FU Consortium
COMMENT MELILO FU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP227, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.gouzy@louloulse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.

FEATURES
Source
1..286550
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
70..813
/gene="gstII OR Smc00916"
70..813
/gene="gstII OR Smc00916"
/EC-number="2.5.1.18"
/function="miscellaneous"
/note="Product confidence : putative
Gene name confidence : putative
Predicted by Codon-usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not experimental
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/transl_table="MADLSFPTTARWPAKNPDIQLYSLPTNGKVSJALBELCIA
YEARRISFDNTEKSPFELSLINPGRIPALIDPGRGKPIGLRESGAILIYLAETG
KLPADARREYELTCWVPMFGIGIFGPFKPADKVNANPYVQRYDDEAKR
LLGVLEARLEBROWLMGGRNTPSPTRPTGKMGKRGSSVAPRCSTIRASRMWMTGI
AAWALRPRRMRKRSRARSRSRSLPRAGD"
588..779
/gene="SMc00915"
588..779
/gene="SMc00915"
/function="miscellaneous: hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon-usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/transl_table="MADGFEYTIADIATYPMVGARKFYGALEVDYKSPFNMDVY
DRGLARPAKGMELPRKE"
867..2084
/gene="SMc00914"
867..2084
/gene="SMc00914"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon-usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11

```



126772 GGTGGCATCATGATGCGGCGGACATGGCGTTCCGGCGCATGAACA 126821  
 131 snleuSerPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSer 147  
 126822 ACATTTCCTTGGGATGATGCGGCGCTCTCTGCTGCTGCTTGG 126871  
 148 LeuPheAlaProGlyLysPglYsnLeuGlySerGlyValGlyTrpAla 164  
 126872 ATGTGTGGAGAGTCTCTGAGGCGCTATGGGCGGCGGCGGCTGGAC 126921  
 164 IleuTrpProLeuSerThrArg...GluAlaGlyTrpSerMetAspL 180  
 126922 GATCTATTCGGCCATCTCGACATCCGCGCATGCCGCGGCGCGCATGATC 126971  
 180 euAlaIlePheAlaValHisLeuSerGlyAlaSerIleMetGlyAla 196  
 126972 TTCGGAATCCCGGCGGATGACATGCGGCGGCGCTGCGATCCCGGCGG 127021  
 197 IleAsnMetIleThrPheLeuAsnMetArgAlaProGlyMetThrIle 213  
 127022 ATCACTTCTACCAACCATCTCAACATCCGCGCTCCGCGCATGACGCT 127071  
 213 unHisValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeu 230  
 127072 GCACAAGATGCGGCTTGGCTGCTGCTGCTGCTGATCACCGCTTCTGCG 127121  
 230 IleLeuLeuAlaLeuProValIleAlaGlyAlaIleThrMetLeuLeuTr 246  
 127122 TCCGCTCTGCGGCGGCTGCGGCGGCGGATGACCATGCTGCTGACG 127171  
 247 AspArgAsnPheGlyThrThrPheAsnProAlaGlyGlyLysAspTr 263  
 127172 GACCGGACATCTCGGCGACGACCTTCTGCGCGGAGGCGGCGGCGACCC 127221  
 263 oIleLeuTrpGlnHisIleLeuTrpPheGlyHisProGlyValTrp 280  
 127222 GATCTGTTCACACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127271  
 280 IleIleIleLeuProGlyPheGlyIleIleSerHisValAlaSerTrpPhe 296  
 127272 TTCGATTCGCGGCGGCTGCGGCGGCGGATGACCATCATCTGACCTTC 127321  
 297 SerLysLysProValPheGlyTrpLeuProMetValTrpAlaMetValAl 313  
 127322 TCGGCGAAGCGCATCTCGGTATCTCGGATGCGCATGCCCATGCGCGC 127371  
 313 aIleGlyValLeuGlyPheValValTrpAlaHisHisMetTrpTrpValG 330  
 127372 GATGCGGCGGCTGCGGCTGCTGCTGCGGCGGCGGATGATGATGATGATG 127421  
 330 LysMetSerLeuThrGlnGlnSerTrpPheMetLeuAlaTrpMetValIle 346  
 127422 GCATGTGCTGCGACAGCGCTACTTCTGCTGCTGCGGACGATGCTATC 127471  
 347 AlaValProThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpG 363  
 127472 GCGGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATG 127521  
 363 yGlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetP 380  
 127522 CGGCTGCGATCGGCTGCGGCGGCGGATGATGATGATGATGATGATGAT 127571  
 380 heLeuPheThrValGlyValThrGlyIleValLeuAlaGlyAlaGly 396  
 127572 TCGTGTACCGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 127621  
 397 LeuAspArgAlaTrpHisAspThrTrpTrpValValAlaHisPheHisTr 413  
 127622 CTCGACCGCTGCGGCGGCGGATGATGATGATGATGATGATGATGATG 127671  
 413 rValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTrpPhe 430  
 127672 CCGTGTGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 127721

430 yrmPheProLysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeu 446  
 127722 GGTTCGCCAAGATGACGCGTACATGATCTGCGAGTTTCTGCGCAACTG 127771  
 447 HisPheTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnH 463  
 127772 CATTTCTGGGATGATGCTGCGGCGGATGATGATGATGATGATGATGATG 127821  
 463 sPheLeuGlyArgGlnGlyMetProArgTrpTrpIleAspTrpProGlu 480  
 127822 TTTCTCGGCGGCTGCGGCGGATGATGATGATGATGATGATGATGATG 127871  
 480 laphAlaLeuTrpAsnLysValSerSerTrpGlyAlaPheAlaPhe 496  
 127872 CCTTTCGCGGCGGCGGATGATGATGATGATGATGATGATGATGATG 127921  
 497 AlaSerPheLeuPhePheIleValIlePheValTrpTrpLeuValAlaG 513  
 127922 GTGCGGCTGCTGATCTTCTCT...TTCTGTGCTGCGGCGGATGATG 127968  
 513 yArgArgGluThrArgProAsnProTrpGlyGluPheAlaAspThrLeuG 530  
 127968 GAAGCGGCTGCGGCGGATGATGATGATGATGATGATGATGATGATG 128018  
 530 lUTrPThrLeuProSerProProAlaHisThrPheGluThrLeuPro 546  
 128018 ATGGCAGAGTGTCTTCTGCGGCGGCTGCGGCGGCTGCGGCGGATGATG 128068  
 547 Lys 547  
 128069 CCG 128071

seq\_name: gb\_ba:R1COXA  
 seq\_documentation\_block: 1707 bp DNA BCT 25-APR-1994  
 LOCUS R1COXA  
 DEFINITION R. leguminosarum coxa gene.  
 ACCESSION X74341  
 VERSION X74341.1 GI:396698  
 KEYWORDS cytochrome c oxidase; cytochrome c oxidase subunit I.  
 SOURCE Rhizobium leguminosarum.  
 ORGANISM Rhizobium leguminosarum.  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Rhizobiaceae; Rhizobium.  
 REFERENCE 1 (bases 1 to 1707)  
 AUTHORS Gabel,C.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUL-1993) C. Gabel, The Johns Hopkins University,  
 Dept. of Biology, Charles and 34th Sts., Baltimore MD 21218, USA  
 REFERENCE 2 (bases 1 to 1707)  
 AUTHORS Gabel,C., Bittinger,M.A. and Maier,R.J.  
 TITLE Cytochrome a3 gene regulation in members of the family  
 Rhizobiaceae: comparison of copper and Rhizobium tropici  
 Bradyrhizobium japonicum and Rhizobium tropici  
 Appl. Environ. Microbiol. 60 (1), 141-148 (1994)  
 9416196  
 FEATURES  
 source  
 1. 1707  
 /organism="Rhizobium leguminosarum"  
 /strain="USDA 27"  
 /db\_xref="taxon:384"  
 91. 1707  
 /gene="coxa"  
 91. 1707  
 /gene="coxa"  
 /EC\_number="1.9.3.1"  
 /EC\_start=1  
 /EC\_end=1  
 /transl\_table=1  
 /product="cytochrome-c oxidase, subunit I"  
 /protein\_id="CAA52388.1"  
 /db\_xref="GI:396698"  
 /db\_xref="SWISS-PROT:Q08855"

gene  
 CDS  
 1. 1707  
 /gene="coxa"  
 91. 1707  
 /EC\_number="1.9.3.1"  
 /EC\_start=1  
 /EC\_end=1  
 /transl\_table=1  
 /product="cytochrome-c oxidase, subunit I"  
 /protein\_id="CAA52388.1"  
 /db\_xref="GI:396698"  
 /db\_xref="SWISS-PROT:Q08855"

```

/translation="MATAAAGENAEDHGHDEHAPPTGRRSTNHKIDIGTLYLIFAI
IAGVIGAAASLAIRALAMYPGEYIEHTLYVFTSYEIVIFEMVPMAMGGGNW
FLPMIGADPMAPRRNNISFWLLPASPGLLMEYIEEVEEPANAGAGMTMYVPLSS
SGHPAPVADLAIFESLIAGASSTIGAINFITITLNNRAGMILHKMDLPEAMSVLTAR
LLLSLPVLGATITMLTDNRNCTTFEAREGGEDLLYOHLEWFGHPEVYLLILPGF
GMHSITSTFSKRPVFCYIGMYAMALIGLGFVVAHMYTVGMDLDEAVFSATM
IIVAPFGIKIEFWIATMWSGSIIEFAPMLMALAFLELFVGGVGVLANSLDRLVH
DYYVVAHMYVLSLGAIFAFAGMYWPKSKCYMNETLEAHFWILFTISVNLIFE
PEHFLGISGMPRIYIDPDAPAGMNLVSSIGSYISGFSVLEFYCYVDAFAKNVPGD
NPMGAGATLEWTLSPSPVHEFEVLPVE"
BASE COUNT      277 a      537 c      479 g      414 t
ORIGIN

alignment_scores:
    Quality: 1873.50      Length:      542
    Ratio: 4.173      Gaps:
Percent Similarity: 82.841      Percent Identity: 64.207

alignment_block:
US-09-712-768-2 x RLCOA ..
Align seg 1/1 to: RLCOA from: 1 to: 1707

7 HisGlyHisAspHisGlyGlyGlyPhePheThrArgTrpPhe 23
|||||
133 CATGCCCATGATGACGACGACCGG.....ACGGCTGGCGCCG 173
23 tSerThrAsnHisLysAspIleGlyLeuLeuTyrrLeuValAlaAlaGly 40
|||||
174 CTCGACCAACCATTAAGATATCGGTACCGTACCTGATCTTCGCTATCA 223
40 aValGlyPheIleSerValLeuPheThrValTyrrMetArgLeuGlyLeu 56
|||||
224 TCGAGAGAGATCGGGGACCATGTCCTGGCCATCCGCGGAACTG 273
57 MetAspProGlyValGlnTyrrMetCysLeuGlyGlyAlaArgLeuIleAl 73
|||||
274 ATGACCCGGGCGGTGAGTACTTC..... 297
73 aAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMetV 90
|||||
298 .....CATACACCCATCTGTACACGATATTCG 325
90 aLhrTyrrHisGlyIleLeuMetMetPhePheValGlyIleProAlaLeu 106
|||||
326 TTACCTCGACAGCGGTTATCATGATCTTCTTCATGTCATGCGGCTATG 375
107 PheGlyGlyPheGlyAsnTyrrLeuMetProLeuGlnIleGlyAlaProAs 123
|||||
376 ATGGGGGGCTTCGGCAACTGTTCCGCCCTCATGATGTCGCCGCCCA 425
123 pMetAlaPheProArgMetAsnAsnLeuSerPheTrpLeuPheIleAlaG 140
|||||
426 TATGGCTTCCCGCGGATGAAATATCTGCTTGGCTCCGCCGCTT 475
140 LyrThrIleMetGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGln 156
|||||
476 CGTTCCGCTGCTGCTCATGTCACCTTCGTCGAAGGGAACCGGGTGC 525
157 LeuGlySerGlyValGlyTrpValLeuTyrrProProLeuSerThrArg.. 172
|||||
526 AATGGCGGGGCTGGCGGAGATGATGTCGCCCTCATCATCAACAG 575
173 GlnAlaGlyTyrrSerMetAspLeuAlaIlePheAlaValHisLeuSerG 189
|||||
576 CCATCCGCGCGCGCGGTGATCGCAATCTTTCGTCGCAATATGCGG 625
189 LysAspSerIleMetGlyAlaIleAsnMetIleThrThrPheLeuAsn 205
|||||
626 GCGCTTCGTCGATCTCGGTGCGATCACTCATCAACAGATCCGCAAC 675
206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSe 222

```

```

|||||
676 ATCGCGGCTCCCGGATGACGCTCCCAAGATGCCCTGTGGCCGGTGC 725
222 rIlePheIleThrAlaTrpLeuIleLeuLeuAlaLeuProValIleAlaG 239
|||||
726 GGTTCGATACCGCATCTCCTGCTGCTTGCACACTCCCGGATTCGCGCG 775
239 LysAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePhe 255
|||||
776 GCCCATTCACCATGCTGTTGACCGACGCAATTCGGCGACGACCTTCTT 825
256 AsnProAlaGlyGlyLysAspProIleLeuTyrrGlnHisIleLeuTrpP 272
|||||
826 GCGCGGGAAGCGCGGCTGATCCGCTGCTTCACACGACCTCTTCGGTT 875
272 ePheGlyHisProGlyValTyrrIleIleIleLeuProGlyPheGlyIleI 289
|||||
876 CTTCGGCATCCGGAAGTCTCATCTCTGATCTCGCCCGGTTCCGCCATGA 925
289 IeSerHisValIleSerThrPheSerLysProValPheGlyTyrrLeu 305
|||||
926 TCAGCCCATCATCTCGACCTTCGCGCAAGCGGATTCGGCTACATC 975
306 ProMetValTyrrAlaMetValAlaIleGlyValLeuGlyPheValValTr 322
|||||
976 GCGATGCTATACGCTATAGCTGCTCTATCCGGGCGCTGGGGTGTGTCTG 1025
322 pAlaHisHisMetTyrrThrValGlyMetSerLeuThrGlnGlnSerTyrr 339
|||||
1026 GCGCGCACCATGATATCTGTCGCGATGATGATGATACCGGAAGGCTATT 1075
339 hMetLeuAlaThrMetValIleAlaValProThrGlyIleLysIlePhe 355
|||||
1076 TCGTTCGCGCGACCATGATCATTTGCTGTGCCAGCGGATCAAGATCTTC 1125
356 SerTrpIleAlaThrMetTrpGlySerValGluPheLysSerProMet 372
|||||
1126 TCATGATATGCGACCATGTGGGTGTGATGATGATGATACCGACCCCAT 1175
372 tLeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyGlyValThrG 389
|||||
1176 GCTGTGGGCGCTCGGGCTTATCTTCTGTTACGCTTGGCGCGCTTACCG 1225
389 LysIleValLeuAlaGlnAlaGlyLeuAspArgAlaTyrrHisAspThrTy 405
|||||
1226 GGGTGTCTCTCGCTAACCGGACGTCGATCGGCTCCACGATACCTTAC 1275
406 TyrValValAlaHisPheHisTyrrValMetSerLeuGlyAlaIlePheAl 422
|||||
1276 TTACGTCGTTGCTCATTTCCACTACGCTCTGTGCGTGGGGCTATTTTCGC 1325
422 aIlePheAlaGlyIleTyrrPheTyrrMetProLysPheSerGlyArgAla 439
|||||
1326 CATCTTCGCGGCTGTACTACTGTGTTCCGAAGATGTCGGGGATATATG 1375
439 hProGlnTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGlyAla 455
|||||
1376 ACAACGAACCCCTGGCGGACGACACTTGTGCTGATCTTCATGCGGCTC 1425
456 AsnValIleThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetPro 472
|||||
1426 AACCTGATCTTCTTCCCGGACACTTCTCGGCAATTCGCGGAAGCCGCG 1475
472 gArgTyrrIleAspTyrrProGlnAlaPheAlaLeuTrpAsnLysValSer 489
|||||
1476 CCGGATATTCAGTACACCGGAGCGGCTTGGCGGCTGGAATCTGTTTCAT 1525
489 eTyrrGlyAlaPheLeuAlaPheAlaSerPheLeuPheIleValIle 505
|||||
1526 CCATCGCGCTGTACATCTCGGGTTCTCGGCTGCTGCTTACTACTGC 1575
506 PheValTyrrThrLeuValAlaGlyArgArgGlnThrArgProAsnProTr 522
|||||

```

```

1576 ...GTCACGATCGCTTGGCCAGAACGTCGCCGAGCGACATCCCTG 1622
522 pclygluPhueAAspThrlengluThrlProserProproPro 539
1623 GGGCGGGGGCGGACACCCTCGAATGAGCGCTCCCTCGCGCGCGG 1672
539 lahlsthrPhggluThrlleuProlys 547
1673 TCACGAGTTCAGACGCTCTGCCCCG 1698

seq_name: gb_ba:AE008010

seq_documentation_block:
LOCUS AE008010 10835 bp DNA BCT 14-AUG-2001
DEFINITION Agrobacterium tumefaciens strain C58 circular chromosome, section
68 of 254 of the complete sequence.
ACCESSION AF008010 AE007869
VERSION AE008010.1 GI:15155740
KEYWORDS
SOURCE
ORGANISM Agrobacterium tumefaciens.
            Agrobacterium tumefaciens
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE
            1 (bases 1 to 10835)
            Hinkle, G., Slater, S.C. and Goodner, B.
            Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
            Unpublished
            2 (bases 1 to 10835)
            Hinkle, G., Slater, S.C. and Goodner, B.
            Direct Submission
            Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
            Cambridge, MA 02139, USA
FEATURES
            location/Qualifiers
            1. 10835
            /organism="Agrobacterium tumefaciens"
            /strain="C58"
            /db_xref="taxon:358"
            complement(118..642)
            /gene="AGR_C_1396"
            complement(118..642)
            /gene="AGR_C_1396"
            /note="(AJ242592) hypothetical protein"
            /codon_start=1
            /transl_table=11
            /product="AGR_C_1396p"
            /protein_id="AAK86575.1"
            /db_xref="GI:15155741"
            /translation="MRLSPARFLAAAGLAIAPAAALAAQOPPGTPTKSTHGAMSV
            /IDKPAASEFOCALMONYIADDPREVGLSVILTKTADRSRLRLAPLGLVKDGM
            ELIYDNNNTGRAVYTRCFSEGGVEVDIDDELTKLVRAGKNVFPALRESVDQDRVGI
            P
            IELSGPASTDLP"
            988..1938
            /gene="AGR_C_1397"
            988..1938
            /gene="AGR_C_1397"
            /note="(AJ242592) cytochrome c oxidase subunit II"
            /codon_start=1
            /transl_table=11
            /product="AGR_C_1397p"
            /protein_id="AAK86576.1"
            /db_xref="GI:15155742"
            /translation="MDSASVYIDDOTREGGVYTRIRIYAALAGMTCLTAVGAHADDP
            VHMOMQEAAPRIMHEISRYEEDYTLMPFLVPLTLLIIVAVKFAKNPVAASKT
            SHNTAIEVWTLPVILFLFLAPPSNLNAOITOPENPDITLKATQIMWSYEKA
            AEGAEPLSPDSYLLKQDRAAKEDKARIPRLIYADNEKVPYGVKTVRLIYTAATD
            VHAFAFAPAGVKIDAVPGRINLWTKRPEKEGGLYGCSELGCKDHAFMPAIKRVSE
            OQYNTMHAASDLSNGANRALMASVDGAPRTVDYVANAETN"
            2549..3664
            /gene="AGR_C_1399"
            2549..3664
            /gene="AGR_C_1399"
            /note="(AJ242592) cytochrome c oxidase subunit III"
            /codon_start=1
            /transl_table=11
            /product="AGR_C_1402p"
            /protein_id="AAK86578.1"
            /db_xref="GI:15155744"
            /translation="MTVIDRDMGAESSELSFAGRDYFELIKPRVMSLVYFPAFAG
            LVAPGEINILGLIILCIAGAGAGCALNMVYDADIDAVMTAKRPIPSGRIAPR
            EIALFGITLFAFVILGLIIVNMFAGLAFTEFAVYVYTMKLRSTQNTVIGGA
            GAPPMGACVAGVSDVILFLIPLMTPAHMALEFRKRDGAAGIIPMPVAV
            GERSTKQMTIVYLVIRAAVAPFTGAGVIAAGVIAFAVYICSPDVRMPEGD
            EKMIPAKMFAVSVLYLFALEGLADHFAIKAVISGL"
            5167..5784
            /gene="AGR_C_1406"
            5167..5784
            /gene="AGR_C_1406"
            /note="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
            /codon_start=1
            /transl_table=11
            /product="AGR_C_1406p"
            /protein_id="AAK86579.1"
            /db_xref="GI:15155745"
            /translation="MANMEATFAGSQRVSNRSILVLCVFPCAMIGMAVAAVPLYS
            LFCRVGNGTQVROYSDVILDKTINVFEDANTSGMLMDRPVDKVKPKIGETI
            KVTAKATRSVATTTGTAVRVPTMEGACFENKVECCPFTTLQPELTLEMVVEFV
            DPDATAREKTSVHTLTSTFPAKTEKVASLPTVTEGDSKL"
            5846..6721
            /gene="AGR_C_1407"
            5846..6721
            /gene="AGR_C_1407"
            /note="(AJ242592) cytochrome c oxidase subunit III"
            /codon_start=1
            /transl_table=11
            /product="AGR_C_1407p"
            /protein_id="AAK86580.1"
            /db_xref="GI:15155746"
            /translation="MADTHOKNHDYHIIDPSMPDLASIGAFIMTFPGGVCMRYLSG
            SFKLFGARLANPMLFETIGIVLYLVYVMAAMADTIRKENEGSGEIVIDMHLPIYMTVL
            ASEVMEFVAMPATFADSLPHEALQASRLAEMVTEGOMPKEIGIEIVDMHLPIYMTVL
            LLSGTCVVAHMLHNDKRLISGLALTVLGYLFSVQYVEYIINAPPEKNSIYGA
            TFFMATIGHGHHFVGVFLLVCLFRAIAGGFTPKOHGFEAAAMVHFDVYVLFLE
            FAIYMGWGAFLNG"
            6941..7369
            /gene="AGR_C_1409"
            6941..7369
            /gene="AGR_C_1409"
            /note="(AJ242592) hypothetical protein"
            /codon_start=1
            /transl_table=11
            /product="AGR_C_1409p"
            /protein_id="AAK86581.1"
            /db_xref="GI:15155747"
            /translation="MYSRFGQREREVLMSEPSNGNEAPVPPIYVGLKGGCPCGNGR
            LFDGFLTPKACGLDYGFDADGDGAVFVMLLVGLVVGMLMWDORFAPVMAH
            VMLMLPFIYVSLVILRLKLGIMIALQYRNNASBGRDRE"

```



```

gene      7338..8114
CDS        /gene="AGR_C_1411"
           /gene="AGR_C_1411"
           /note="(AF182952) Surfelt 1"
           /codon_start=1
           /transl_table=11
           /product="AGR_C_1411p"
           /protein_id="AAK86582.1"
           /db_xref="GI:15155748"
           /translation="MQAKAGLVNDVSHMKAOBRVMPAPLVLAIALILGCTWQY
           KRVKELMADIEERNASPATLSDEIAIVKSGEIEYKVRILSGTFDHTREHFA
           THQGTGYITPTPLADGRILFVNRGFVEPMKEAARPDGQVSGEVTITGLARAPL
           VAKBSLLPDDNDIAKNIFYMKDLAMASSAEVPEERLVKLFVDADNAPNGGMPQGGV
           TLIDLPNNHLQYAITWGLAALVIVAGFAFRNGKAGE"
gene      8173..9219
CDS        /gene="AGR_C_1414"
           /gene="AGR_C_1414"
           /note="(AE006203) LyCB"
           /codon_start=1
           /transl_table=11
           /product="AGR_C_1414p"
           /protein_id="AAK86583.1"
           /db_xref="GI:15155749"
           /translation="MNTAVSKRPPLTILCGPRGCGAVDRAIQIVLALAKAYAPVY
           NHEIVHNRYVEGLEAKGALFVEELHEIPAEHRQPVVFAHGPVKSVDPDAQANLF
           YLDATCPLEKSVNHOAMHQRHVVILIGHAGHEVIGTMGOLPEGTVALREVEDA
           GVEBPVRENLFVQTOTLLSYDPAVYIARLOERFPALQAPADQATNTODADK
           QAAGCDLFIYVGAIPNSNSKRLVEVALRAGAKRIVVORASISDMETIGDITVGS
           AGASPEVIVDEITIEAFKARPDITLDLAVVEIEHFLVNRRLSIELITDDMAFVNG
           NASNALPPKATAP"
gene      9312..10280
CDS        /gene="AGR_C_1416"
           /gene="AGR_C_1416"
           /note="(HOMOSERINE KINASE (HK) )"
           /codon_start=1
           /transl_table=11
           /product="AGR_C_1416p"
           /protein_id="AAK86584.1"
           /db_xref="GI:15155750"
           /translation="MAVYTDITDEDELNPLTOYDVGSLYSKGIARGVENSFLHTT
           KDLITLYEKREKNDLPFELGIMOHIAKGSCLPLPRKGGELIGELSGPALLI
           SPLEGMWLRPEAKHCEVKAALAAHMLASGEFETRPNALSYDGMKVLMDKSEERAD
           EYKGLREETRPETDILAHMPKDLPGAVIHADLFQDNVEFLDELGLIDFEYFACND
           LLAIVDSICLNAMCFEKDGAIVNTKGALLEGQSVRPSEAELEALPLISRSALRF
           FLRLYDMVLTTPAGALVKKDPLEYLRKLFHRTIANVAEYGLAGE"
gene      10277..10717
CDS        /gene="AGR_C_1417"
           /gene="AGR_C_1417"
           /note="Ribonuclease H (E.C.3.1.26.4) Mutant with His 62
           Replaced by Pro (H62p)"
           /codon_start=1
           /transl_table=11
           /product="AGR_C_1417p"
           /protein_id="AAK86585.1"
           /db_xref="GI:15155751"
           /translation="MKHVDLFTDGACSGNPGQGWAVLRGTEKEKLSGSEADTTNN
           RMELLAISALNALKSPCEVDLYTDSAYVKGDTIKWIFGMKKGMWKTADNKPVKWVEL
           MWLLEAEOEHKVTLLHWKGHAGHPENERADELARKGMEPFKRR"
BASE COUNT 2021 a 3194 c 3242 g 2378 t
ORIGIN
alignment_scores:
  Quality: 1872.50      Length: 548
  Ratio: 4.018          Gaps: 6
Percent Similarity: 85.036 Percent Identity: 63.869
alignment_block:
us-09-712-768-2 x AE008010

```

```

Align seg 1/1 to: AE008010 from: 1 to: 10835
7 HISGLYHISASPHISHSISGLULYSGLINGLYPHEPHEHTRARGTPHPHME 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2031 CATTCGCATGATCATTTCCATTAAGCCGGGCTTTTGGCCCGCTGGTTCT 2080
23 tserthrashnshislyasphileglyleuleuthrleuvalalaglylv 40
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2081 GTCCACCAACCAACCAAGGACATCGCCACCTGTACCTGATCTTCGCCATCA 2130
40 alvaliglypheelleservalleuphehtrvaltyrmetargleuleu 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2131 TGGCGGATATCATTCGGCGCGGCTTTGGTCGTCATCGATGAGAGCTG 2180
57 Metaspproglyvaligntrmetcysleuigngly...Alargleu1 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2181 CAGGAGCGCGGCATCCAG.....ATCTTCCACGGTCTGGCTTCATGCT 2224
72 eAlasphalaserglnthrlysthrAlaansgly.....HisleutrpA 87
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2225 CTACGGCTTCGAGCGGATGCGCCATCGACGGGATGAAGACATGTTCA 2274
87 snvalmetvalthrtyrhisglylleuumeetmetphehtrvalglyle 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2275 ACGTGTTCACGACACGCGCGGCTGATCATGATCTTCTTATGATGTCATG 2324
104 ProhaleupheglyglypheglyAsnTyrluumeetProleuign1leg1 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2325 CCGGAGATGATCGGGGTTTGGCAACATGATGATCCCGATCATGATGATCGG 2374
120 yAlaProasphalaaphneProaGmetasnasleuSerPhehtrleup 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2375 CGCGCTTACATGCTTTCCCGCGCCT.AMACATATTTCTTCTTGCTGCTCA 2423
137 heilealaglyThrAlamerclyValAlaserleuphealaproglygly 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2424 TCGTTCGGCGCTTCATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2473
154 AspAlaGlnleuGlyserglyValGlyTrpValleuhyproproleuse 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2474 GCGGCTGCTTATGCGCGTGGTGGCGGCTGACATGATATCGGCTTCTGTC 2523
170 trhArg...GluAlaglytyrsermetaspleuAlailepheaValh 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2524 GACAGATGACATGCCCGCGCGCGGCGGTGATCTTCGATCTTCTGCTGC 2573
186 IsteuSerglyAlaserSerilemetcylalaleasmetilehtrThr 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2574 ACGTCGCGCGGTGATCCTCGATCCTCGGCCCATCAACTTCATCACACAC 2623
203 PheleuasmetatargalaproglymetThrleuHislyValproleup 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2624 ATCTGTGAACATGCCCGCGCGGCGGATGACCTGACACAGATGCCGCTGTT 2673
219 eSerTTPSerilepheelThrAlatrleuileuleuleuAlaaleuprov 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2674 TGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2723
236 alleuAlaglyAlailehtrmetleuuthrAsparGasnphleglythr 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2724 TTTTGGAGGTGATCATCACCATCTTCTGACCGACCGTAACTTCGCGCAG 2773
253 ThrPhePheasnProalaglyglylyAspproileuTyrcinHis1 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2774 GCGTCTTCTCGCGCGGAGGCGGCGGCTGATCTGCTTACACGATCT 2823
269 eleutrphepheglyHisProgluValtyrleileileuleuProglyp 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2824 GTTCGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2873
286 heglyleileleserHisValaserThrPheSerlystysProvalp 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```







```

1502 GCTATACGCGCTGTCACCGCCGACGCTGATCATGATCTTCTTCATG 1453
102 GYLILEPROALALEuPheGlyGlyPheGlyAsnTYrLeuMetProLeuG1 118
1452 GTCATGCTGCGATGATGCGGGGTTGCGCAACTGTTGCTTCGATCAT 1403
118 nllleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeuSerPheT 135
1402 GATCGGCGCGCGGACATGCGCTTCCCGGATGAGAACACATCTGCTCT 1353
135 rPhePheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAlaPro 151
1352 GGTGCTGATCGCGCGCTGATCTCTGCTGCTGCTGCTGCTGCTGCTG 1303
152 GYLILYAspGlyGlnLeuGlySerGlyValGlyTYrValLeuTYrPro 168
1302 GCGCGCGCGCGCGCAC...GGCTTGGCGGGGTGGTGGACGATCATCGCG 1256
168 oleuSerThr..ArgGluAlaGlyTYrSerMetAspLeuAlaIlePheA 184
1255 GCTGTGACCATCGGTACAAAGGTGCGGCGCATGACATGCGCATCTGT 1206
184 lAlaAlaHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIle 200
1205 GGTTCACCTGCGCGCGCGCTGCTGCTGCTGCGCGCGCATCACTTCAT 1156
201 ThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisLysValPr 217
1155 ACCACGATCTTCATATAGCGCGCGCGCGCATGACCTTGCACCGCATGCG 1106
217 oleuPheSerThrSerIlePheIleThrAlaTrpLeuIleLeuLeuAlu 234
1105 GCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1056
234 euProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPh 250
1055 TGGCGCGTGGCGCGCGCGCATGACCATGCTGCTGACCGGACGTAAGTTC 1006
251 GYLIThrThrPhePheAsnProAlaGlyGlyLysAspProIleLeuTYrG1 267
1005 GGCACCGCATCTTCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 956
267 nHisIleLeuTrpPhePheGlyHisProGluValTYrIleIleIleLeuP 284
955 GCACCTTGTGCTGCTTTCGCGCACCGGAAAGTATATCTGATCTGCTGC 906
284 roGlyPheGlyIleIleSerHisValValSerThrPheSerLysLysPro 300
905 CGGCGTTGGCATCATGACGACATCTGCTGACCTTCTCCAGAAAGCGCG 856
301 ValPheGlyTYrLeuProMetValTYrAlaMetValAlaIleGlyValIle 317
855 GTCTTGGCTACCTGCGGATGCGCTACGCGATGGTCCGCTTCGCTTCGT 806
317 uGlyPheValValTYrAlaHisHisMetTYrThrValGlyMetSerLeuT 334
805 CGGCTTGTCTGCTGCGCGCACCACTACAGCGTGGCATGACATCA 756
334 hrgInGlnSerTYrPheMetLeuAlaThrMetValIleAlaValProThr 350
755 ACCTGCGCGCTACTCTGCTGCGCGCACGATGATCATCGCGCTCCGAC 706
351 GYLILYsIlePheSerTrpIleAlaThrMetTrpGlySerValG1 367
705 GCGCTGAGAGATCTTCACCTGGATGCGACACATGTTGGCGGTTGCGTCA 656
367 uPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPheThrV 384
655 CTTCAGAGCGCCATGCTGTTGGCGGATCGGCTTATCTTCTGTTCAAG 606
384 aLgLYGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAspArgAla 400

```

```

605 TCGGCGCGCTGACCGCGCGCTGCTGTCGAAACGCGCGCATGACAC 556
401 TYrHisAspThrTYrTYrValValAlaHisPheHisTYrValAlaMetSerIle 417
555 CTGCGACGACCTACTAGCTGAGTGGCGCGCATTCGACATGACGCTGCTGCT 506
417 uGlyAlaIlePheAlaIlePheAlaGlyIleTYrPheTYrMetProLysP 434
505 GCGCGCGGTGTTCCGCTTCTGCGCGCTTCTACTAGCTGCTGAGAA 456
434 hSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTrpThr 450
455 TGTGGCGGTGAGTACAAACAGTTCCTGCTGCGCGCGCATCTGCTGCTG 406
451 PhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuGlyArg 467
405 AAGTTCATCGGTGAGTACCTGATCTTCTGCGGACGCTTCTGCTGCTG 356
467 gLInGlyMetProArgArgTYrIleAspTYrProGluAlaPheAlaLeuT 484
355 GCAAGGATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
484 rPAsnLysValSerSerTYrGlyAlaPheLeuAlaPheAlaSerPheLeu 500
305 GCACTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
501 PhePheIleValIlePheValTYrThrLeuValAlaGlyArgArgTYrTh 517
255 GGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 209
517 rArgProAsnProTYrGlyLysPheAlaAspThrLeuGluTrpThrLeuP 534
208 CGAGCGCAACCGCTGCGGAGGCGCGCACCGAGCTGAGTACTGCTGCT 159
534 roSerProProAlaHisThrPheGluThrLeuProLysArgSerLysP 550
158 CCGCGCGCGCGCGCTTCCACGATTCAGCAACGCGCGTATCAAGGCG 109
551 TrpAspLysHis 554
108 GACGACCGACAC 97

seq. name: gb_AP03011
seq. documentation block:
LOCUS AP03011 346510 bp DNA BCT 15-MAY-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 18/21.
ACCESSION AP03011 BA000012
VERSION AP03011.2 GI:14026664
KEYWORDS
SOURCE
ORGANISM
Mesorhizobium loti (strain:MAF303099) DNA.
Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Ido, S., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsunoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpoto, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
JOURNAL
MEDLINE
21082930
DNA Res. 7 (6), 331-338 (2000)
AUTHORS
Kaneko, T.
TITLE
Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/rhizobase/
Tel: 81-436-52-3935 (ex. 2338), Fax: 81-436-52-3934)
JOURNAL

```



```

CDS
    /gene="mlr7098"
    9079..10692
    /gene="mlr7098"
    /codon_start=1
    /transl_table=11
    /product="GMC-type oxidoreductase"
    /protein_id="BAB53270.1"
    /db_xref="GI:14026674"
    /translation="MSETSPFGDYIVVAGSAGCYLANRLSNDRNRVLLIAGGS
    DRHWDIPILGILFCMGNPRTDMMKTEPEAGLNSLTPRGKVLGSSVMGMYM
    RGGADADYGRKAGNMGSMDDVLPYRKSSEDYHGKSAHSGGEMRVERORLSWPT
    LDAERDAEELGIEPRIGDFTNGTMEGSGYEENVRGLRPNNTSKAFIRGIAAQNLRV
    VTGARETERLEPDRVYGLLEPGDRLHRSVAGOTILAAQINSPLILEISGVGRPL
    LSTGVDVRHALLPVGNILODHIQIRTVFVANAATLNERHNLSRAAMGVETALKR
    SGPLSMAPSQLGTRASDARLATPDELYHNQPLSTDRGERLHRFPATVYSCNLRPE
    SRGTVHMASSDPREARKIRREYLTSTADRSVAASLRHVSLSMKARAVAFAPENMLP
    GTQYDSEDLIRKAGDITGTTLEFVGTCKMGSDMAVVDRLRVHGLARLVVDASTM
    PTIVSGNTSPVVMIAEKAAEMILLDPR"
    10929..11801
    /gene="mlr7099"
    10929..11801
    /gene="mlr7099"
    /codon_start=1
    /transl_table=11
    /product="esterase"
    /protein_id="BAB53271.1"

alignment_scores:
    Quality: 1846.50      Length: 555
    Ratio: 3.937          Gaps: 8
    Percent Similarity: 84.505    Percent Identity: 61.441

alignment_block:
    US-09-712-768-2 x AP003011 ..

Align seg 1/1 to: AP003011 from: 1 to: 346510
    1 MetAlaSpAlaAlaIleHh...GLYHIsAspHisHisGluIuyscIngl 16
      |||||
    337162 ATGGCAGACCTGCTGACGTCACGACGCGCAGACCAAGCCTTATCATG 337211
      |||||
    16 yPhePheThArgTrpPheMetSerThraSnHisLysAspIleGlyLeul 33
      |||||
    337212 C...TGGGTCGCGTGGGTCTATTGACCAACCAAGACATCGGACGCG 337258
      |||||
    33  euryrLeuValAlaAlaGlyValValGlyPheIleSerValLeuPheThr 49
      |||||
    337259 TCYACCTGATCTTCGCGATCATGGCGGCATCGGCGCGCTGTGCG 337308
      |||||
    50 ValTyzMetArgLeuGluLeuMetLaspProGlyValGlnTyzMetCysLe 66
      |||||
    337309 GTCCGCATCCGCATGAGTTCGACGACCGCGCATCCAG...ATCTT 337352
      |||||
    16 yPhePheThArgTrpPheMetSerThraSnHisLysAspIleGlyLeul 33
      |||||
    337212 C...TGGGTCGCGTGGGTCTATTGACCAACCAAGACATCGGACGCG 337258
      |||||
    33  euryrLeuValAlaAlaGlyValValGlyPheIleSerValLeuPheThr 49
      |||||
    337259 TCYACCTGATCTTCGCGATCATGGCGGCATCGGCGCGCTGTGCG 337308
      |||||
    50 ValTyzMetArgLeuGluLeuMetLaspProGlyValGlnTyzMetCysLe 66
      |||||
    337309 GTCCGCATCCGCATGAGTTCGACGACCGCGCATCCAG...ATCTT 337352
      |||||
    66 uGluGly...AlaArgLeuIleAlaAspAlaSerGlnThrcysThrAla 82
      |||||
    337353 CAGCGGTCGTGGCGAGATGCTTACGCGCATGACGCGGCGCGCATTTG 337402
      |||||
    82 snGly.....HisLeuTrpAsnValMetValThrTyzHisGlyIleLeu 96
      |||||
    337403 ATGGCGGCAAGACGATGATACATGCTTCGCGACCGCGCATGCGCTATC 337452
      |||||
    97 MetMetPhePheValGlyIleProAlaLeuPheGlyIlePheGlyAsnTy 113
      |||||
    337453 ATGATCTTCTTCATGTCATCGCGCGTGTATCGCGGCTTCGCGCACTG 337502
      |||||
    113 rLeuMetProLeuGlnIleGlyAlaProAspMetAlaPheProArgMeta 130
      |||||
    337503 GATGGTGCGCATGATGATGCGCGCGCATGCGCTTCGCGCGCATGA 337552
      |||||
    130 snAsnLeuSerPheTrpLeuPheIleAlaGlyIleThraIleMetGlyValAla 146
      |||||
    337553 ACAACATCTCTCTGCTGCTGCGCGCGCGCTTCATCTGCTGCTGACG 337602

```

```

147 SerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTr 163
    |||||
337603 TCGATGTCCTGTCGCGGAGTGGCGCGCTTCGCGCGCGCGCGCTG 337652
    |||||
163 pValLeuTyTrProProLeuSerThraG...GlnAlaGlyTyzSerMet 179
    |||||
337653 GACGCTCATTCGCGCGCTGTCGACCTCGGTCGACCGCGGCGCATG 337702
    |||||
179 splLeuAlaIlePheAlaValHisLeuSerGlyAlaSerIleMetGly 195
    |||||
337703 ATCTGGCGCATCTGTCGATCCACATTCGCGGCTGCTGCTGATCTGCG 337752
    |||||
196 AlaIleAsnMetIleThrThrPheLeuAsnMetAlaGlnProGlyMetTh 212
    |||||
337753 GCCATCAACTTCATCACACCATCATCAACATGCGCGCTCGCGCATGAC 337802
    |||||
212 rLeuHisLysValProLeuPheSerTrpSerIlePheIleThrAlaTrpL 229
    |||||
337803 GCTGCACAAGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 337852
    |||||
229 euIleLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeu 245
    |||||
337853 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 337902
    |||||
246 ThrAspArgAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAs 262
    |||||
337903 ACCGACCGCAATTCGCGCACCATCTTCGCGCGCGCGCGCGCGCG 337952
    |||||
262 pProIleLeuTyTrGlnHisIleLeuTrpPhePheGlyHisProGlyVal 279
    |||||
337953 TCCGCTGCTGTCACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 338002
    |||||
279 yIleIleIleLeuProGlyPheGlyIleIleSerHisValAlaSerTh 295
    |||||
338003 ACATCTGATCTCTGCGCGGCTTCGCGCATCATGACGCACATGCTCGAC 338052
    |||||
296 PheSerLysLysProValPheGlyTyTrLeuProMetValTyTrAlaMet 312
    |||||
338053 TTCTCGAAGAACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 338102
    |||||
312 lAlaIleGlyValLeuGlyPheValValTrpAlaHisMetTyTrThv 329
    |||||
338103 CCGCATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 338152
    |||||
329 aGlyMetSerLeuThrGlnGlnInsTyTrPheMetLeuAlaThrMetVal 345
    |||||
338153 CCGGCTGTCGCTGCGACACGCGGCTACTGCTGCTGCGCACATGCTG 338202
    |||||
346 lIleAlaValProThGlyIleLysIlePheSerTrpIleAlaThrMetTr 362
    |||||
338203 ATCGCGGTCGCGCGCGGCGGGAAGATCTTCCTGATCGGACGATGTG 338252
    |||||
362 pGlyGlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPhe 379
    |||||
338253 GGGCGGCTGATCTGCTTCAACAGCGCGATGCTGCGGCGCTGCA 338302
    |||||
379 ePheLeuPheThrValGlyGlyValThrGlyIleValLeuAlaGlnAla 395
    |||||
338303 TCTTCTGTCACCATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 338352
    |||||
396 GlyLeuAspArgAlaTyTrHisAspTrpTyTrTyValValAlaHisPhe 412
    |||||
338353 GGGCTGACGCGCGCTGCTGATGACACCTATTCGTATGATGCGGCACTTCA 338402
    |||||
412 sTyValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyTr 429
    |||||
338403 CTACGTGCTGCTGCTGCGCGGCTGCTGCGCATCTTCGCGGCTGCTACT 338452
    |||||
429 heTyzMetProLysPheSerGlyArgAlaPheProGluTrpAlaAlaLys 445
    |||||
338453 ACTGATGTCGCGGACGATGACGCGGTACATGATGATGCGGCGGCTGACG 338502
    |||||
446 LeuHisPheTrpThrPhePheIleGlyAlaAsnValThrPhePheProG 462

```

```

|||||.....|.....|.....|.....|.....|
338503 ACCACCTTCGCGACCTTGTGCGGCTCAACCTGATCTTCCCGCA 338552
462 nHisPheLeuGlyArgIngluMetProArgArgTyrIleAspTyrProG 479
|||||.....|.....|.....|.....|.....|
338553 GCATTTCCTTGCGCTCGCGGCGCATGCCGCGGCGGACCATGACATATCCGG 338602
479 LuAlaPheAlaLeuTyrPAsnLysValSerSerTyrGlyAlaPheLeuAla 495
|||||.....|.....|.....|.....|.....|
338603 AGCGGTTTGGCGCGGACACTGCTGCTGCATGCGCTCTATATCTCG 338652
496 PheAlaSerPheLeuPhePheIleValIlePheAlaTyrThrLeuValAl 512
|||||.....|.....|.....|.....|.....|
338653 GCCCTCGGC.....GTGGGATCTCTCTCTACGCGCTGTTCGA 338690
512 aglyArgArgGluThrArg.....ProAsnProTyrGlyAlaPhe 526
|||||.....|.....|.....|.....|.....|
338691 AGCTTCCAGAGAGAGCGCATCCCGGCTGCCACCCATGGGCGAGGCGG 338740
526 IaAspThrLeuGluTyrThrLeuProSerProProAlaHisThrPhe 542
|||||.....|.....|.....|.....|.....|
338741 CCACACGCTCGATGATGCGAGCTGCTTCCGCGCGCTCCACCACTGG 338790
543 GluThrLeuProLys 547
|||||.....|.....|.....|.....|.....|
338791 GAGCAGTTCGCCGAG 338805

seq_name: gb_ba:NMCOXABC

seq_documentation_block: 4380 bp DNA BCT 27-JUL-2001
LOCUS NMCOXABC
DEFINITION N.winogradskyi DNA for coxa, coxB and coxC genes.
ACCESSION X89566
VERSION X89566.1 GI:1177572
KEYWORDS cox gene; coxB gene; coxC gene; cytochrome c oxidase subunit I; cytochrome c oxidase subunit II; Heme O synthase.
SOURCE Nitrobacter winogradskyi.
ORGANISM Nitrobacter winogradskyi.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Nitrobacteraceae; Nitrobacter.
REFERENCE
AUTHORS Berden,G.
TITLE 1 (bases 1 to 4380)
JOURNAL Nitrobacter winogradskyi cytochrome c oxidase genes are organized in a repeated gene cluster
AUTHORS Antonie van Leeuwenhoek. 69 (4), 305-315 (1996)
MEDLINE 96433441
REFERENCE
AUTHORS Berden,G.P.R.
TITLE 2 (bases 1 to 4380)
JOURNAL Direct Submission
Submitted (12-JUL-1995) G.P.R. Berden, Centre de Recherches Agronomiques de l'Estat, Station de Chimie/Physique Agric, Chaussée de Wavre, 115, B-5030 Gembloux, Belgium
FEATURES
Source
1..4380
/organism="Nitrobacter winogradskyi"
/strain="ATCC 14123"
/db_xref="taxon:913"
/clone="plasmid Nw24.92"
/clone_id="CA61743.1"
/clone_lib="Library of G.Berden"
727..1584
/gene="coxA"
727..1584
/gene="coxB"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=1
/product="cytochrome-c oxidase subunit I"
/protein_id="CA61743.1"
/db_xref="GI:1177573"
/db_xref="SPTREMBL:059630"
/transl_table="MKSMGOYGRGRFLGAFGGAFAAGAAFAADEIGOPHAM
GFDLOGVTFVMSIARMNGGLVITLITLFTALLVAVAFNARNRNVPSKPTTHN
TMIEVAMTILPILVIGIAVPSFRLLEQDLDPKADULTYKATCKQWYWTYTPDNCKF

```

```

EFSLDADKKPRLAVDNELVVPVNVAVVIGADVTHSPGVSPGMKVDA IPRGL
NETFKATKTCIGIFGCGSELGGRDHAFMLAVRVSRQEPAAVETAKKFAASNPANA
VASAGAYWO"
727..828
/gene="coxA"
1682..3301
/gene="coxB"
1682..3301
/gene="coxB"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=1
/product="cytochrome-c oxidase subunit I"
/protein_id="CA61744.1"
/db_xref="GI:1177574"
/db_xref="SPTREMBL:059631"
/transl_table="MATASADHADHDGDNHNPYGMKRFATSTNKHGIGTVLVA
IVAGITGLMSMGIRAEIAMPQIYDPDPAHYVVPATNGLIMIPEWMPAMIGFGN
WFPVLMIGADMAEPFRNNISFWLDPAAFAILLISSFMSGEPATGAGCTLYAPLS
TSGHPGPAVDFSTLHLAGSSILGAINRTYIFMRAPGMTHMKPLFVNSIIVTV
FLILALPVYAGATIMLIDRNRTGTPRADGGDVLLOHLFWFGRHEVITLILPG
FGHISHTSIFSKPIETGLMAVAVAGIGIGVVAHNMHTVGNVSAQAYPVAAAT
MTAVPRTGVAFVSMIAIMWGSIEFTPLMFAIGTIFLFTVGVTGVVLANGVDRVM
QETTYVIAHFHYVWGIAVAFIFSGWYFPEKMSGYNETIGLHFWMFIGANTLF
EPQHLGLQGMRRNVDPDAFAGNNEISYGAFTIGFVLIIFLGLVDAFARKQOAA
DNPMGEGATLEWTLSSPPPHQFTLPKVQ"
3314..4303
/gene="coxC"
3314..4303
/gene="coxC"
/codon_start=1
/transl_table=1
/product="Heme O synthase"
/protein_id="CA61745.1"
/db_xref="GI:1177575"
/db_xref="SPTREMBL:051322"
/transl_table="MCAASCAAHPTKAGSIVVDKAVELPRPISSEAGVADYFALL
KPRVMSIVFTALVGMIAFGHVPVGLTILICIAVAGAGALMALLEGIDVIMS
RTANRP1IPRIRIGEMGEGLTISFSGVITGLVNWVAGGLAFTFEFVYITWMS
LKRTANONIVIGAGALPPVYVMAAASCTSEVERLPLLEIFPTPTPTALAFRS
DOYARAGVMP1PVYAGPDRATROILTYTILVLAIAAPLPDIFPMITGVISLIGAG
MLVLAIEVYRHRGSQLRATRKLEFESFYLFAVLDDVAVAKAAPLIM"

BASE COUNT 749 a 1252 c 1415 g 964 t
ORIGIN

alignment_scores:
Quality: 1638.00 Length: 558
Ratio: 4.022 Gaps: 7
Percent Similarity: 81.900 Percent Identity: 61.828

alignment_block:
US-09-712-768-2 x NMCOXABC ..
Align seg 1/1 to: NMCOXABC from: 1 to: 4380
1 MetaLaSpAlaAlaIleHisGlyHisAspHis..... 11
|||||.....|.....|.....|.....|.....|
1682 ATGCGACACGATCTGCTGCAGCAGCATGCCATGACATGATGATGACCA 1731
12 ....HisGluLysGlnGlyPhePheThrArgTyrPheMetSerThrAsnH 27
|||||.....|.....|.....|.....|.....|
1732 TGGCGACCCACCGACAGCG...TGAAGCGCTTCTGCTATTCGACCAATC 1778
27 IsLysAspIleGlyLeuLeuTyrLeuValAlaIleGlyValIleGlyPhe 43
|||||.....|.....|.....|.....|.....|
1779 ACAAGGACATGGGACCATGATCTGCTGCGCATGCGGCGGATC 1828
44 IleSerValLeuPheThrValTyrMetArgLeuGluLeuMetAspProG1 60
|||||.....|.....|.....|.....|.....|
1829 ATCGCGCGCGCTGATGATGCGGCGGCAACTGATGTACCCCGG 1878
60 yValGlnTyrMetCysLeuGluGluAlaArgLeuIleAlaAspAlaSerG 77

```

```

1879 CATACAG.....GTCTTGTGATCCGCG. 1900
77 LnthrcysThrAlaAsnGlyHisLeuTrpAsnValMetValThrTrpHis 93
1901 .....CATACCTACACGCTGTTTGCAGCACATTCAC 1930
94 GlyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGlyGlyPh 110
1931 GGCTCATATCATATCTTCTTCATGGTGAATGCCCTGCATGATCGCGCGCTT 1980
110 eGlyAsnTrpLeuMetProLeuGlnIleGlyAlaProAspMetAlaPheP 127
1981 CGGCACACTGTTGCTGCGCTGATGATCGCGCGCGCGACATGCGGCTTCC 2030
127 roArgMetAsnLeuSerPheTrpLeuPheAlaIleThrAlaMet 143
2031 CGGCATGTGAACACATCTGTTCTGGCTGCTGCGCGCGGCTTGGCGTGG 2080
144 GlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySerG 160
2081 CTGCTATATCTCTCTCTTCATGGAGGTTGAGCCGCGCGCGACCGCGAGG 2130
160 yValGlyTrpValLeuTrpProLeuSerThrArg...GluAlaGlyT 176
2131 CACGGGCTGGACGCTGATGCGCGCGCTGACAGCGAGCGGTATCCGCGT 2180
176 ySerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAlaSer 192
2181 CGCGGGTGGCTTCACATCTGTCATCCACCTTCGCGCGCGCTCTCTCG 2230
193 IleMetGlyAlaIleAsnMetIleThrPheLeuAsnMetArgAlaP 209
2231 ATCCTCGGGCGCATCACTTCATACACGATCTTCACATGCGCGCGC 2280
209 OGlyMetThrLeuHisLysValProLeuPheSerTrpSerIlePheIleT 226
2281 TGGCATGACCATGCACAGATGCGCGCTGTCGTCGTCGATCCTGTGCA 2330
226 hrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAlaGlyAlaIleTr 242
2331 CGGTGTTTGGTGTGCTGCTGCGCGCTGCGGTTCTGCGCGCGCGATCAG 2380
243 MetLeuLeuThrAspArgAsnPheGlyThrThrPhePheAsnProAlaG 259
2381 ATGCTGTGACGACCGGTAATTGCGACACAGCTTCTCGCGCGCATG 2430
259 yGlyGlyAspProIleLeuTrpGlnHisIleLeuTrpPhePheGlyHisP 276
2431 CGCGCGGATCCGCTGCTGTCAGCATGCTGTCGTTCTTCTGCGTCA 2480
276 roGlyValTrpIleIleLeuProGlyPheGlyIleIleSerHisVal 292
2481 CGGAGGGTATCATCGATTCGCGCGCTTCCGATGATCAGCCATG 2530
293 ValSerThrPheSerLysProValPheGlyTrpLeuProMetValTyr 309
2531 ATCTCGACGTTTTCGAAGAGCGGATCTTGGATCTTGGGAGTGGCTA 2580
309 rAlaMetValAlaIleGlyValLeuGlyPheValValTrpAlaHisHisM 326
2581 CGGCATGTCGTCATCGCGCATCGCTTGTCTGTCGCGCGCACACACA 2630
326 eTrpTrpValGlyMetSerLeuThrGlnInSerTrpPheMetLeuAla 342
2631 TGTACAGGTCGTCATGTCCTTCCGCGCATGCTACTTCTGTCGCGGCG 2680
343 ThrMetValIleAlaValProThrGlyIleLysIlePheSerTrpIleAl 359
2681 ACCATGTGTATTCGCGGTCGCGAGCGGCTGAAGTGTTCGTCGATCGC 2730
359 arhMetTrpGlyLysSerValGluPheLysSerProMetLeuTrpAlaP 376

```

```

2731 CACGATGTGGGCGGCTCGATCGAGTTCACAGCCCGCATGCTGTTCCGA 2780
376 heGlyPheMetPheLeuPheThrValGlyGlyValIleThrGlyIleValLeu 392
2781 TCGGCTTATCTCTCTGTTACAGGTGCGGCGCGTGCAGCGGCTGTGCTG 2830
393 AlaGlnAlaGlyLeuAspArgAlaIleThrHisAspThrTrpValValAl 409
2831 GCCATGCGGCGGCTGATGCGGTGATGCAGAGACACTACATCGATCGAC 2880
409 ahISpHeHisTrpValMetSerLeuGlyAlaIlePheAlaIlePheAlaG 426
2881 GCACCTTCACACTACGATGATGGCATTTGCGCGCTGTCGCGATCTTCG 2930
426 yLleTrpPheTrpMetProLysPheSerGlyArgAlaPheProGluTrp 442
2931 GCTGTACTACTGTTCCCGAAGATGTCGGCTACATGTACACAGACACC 2980
443 AlaAlaLysLeuHisPheTrpThrPhePheIleGlyAlaAsnValIleThrP 459
2981 ATCGGCAAGCTGCACATCTCTGGGTGATGTTCAATCGCGCGCATCTG 3030
459 ePheProGlnHisPheLeuGlyArgGlnGlyMetProArgArgTrpIleAl 476
3031 CTTCGCCGACGACACTTCCTGCTGTCAGGCGATCATGCGGCGCATG 3080
476 sPTyrProGluAlaPheAlaLeuTrpAsnLysValSerSerTrpGlyAla 492
3081 ACTATCCGATGCTGTCGCGGCTGGAACGAGATCTGCTTACGCGCGC 3130
493 pheLeuAlaPheAlaSerPheLeuPheIleValIlePheValIleTrpTh 509
3131 TTCATCACGAGGC.....TTCGCGCTCATCATCTTCTCTACG 3168
509 rLeuVal.....AlaGlyArgArgGluThrArgProAsnProTrpG 523
3169 CTTGCTCGACGCTTTCGCGCGACACAGCGGCTGCTACATCCGTGG 3218
523 LyGluPheAlaAspThrLeuGluTrpThrLeuProSerProProAla 539
3219 GTGAGGCGCGGACGACGCTGATGATGACGCTGCTACACGCTTCCG 3268
540 HisThrPheGluThrLeuProLys 547
3269 CACCACTTCTCCACCTGCGGAG 3292

seq_name: gb_ba:BCOXAG

seq_documentation_block:
LOCUS      BCOXAG               1830 bp      DNA          BCT          12-SEP-1993
DEFINITION B.japonicum coxa gene for cytochrome aa(3) subunit I (BC 1.9.3.1).
ACCESSION  X54318
VERSION    X54318.1 GI:39505
KEYWORDS   coxa gene; cytochrome aa(3) subunit I.
SOURCE     Bradyrhizobium japonicum.
ORGANISM   Bradyrhizobium japonicum
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Bradyrhizobium group; Bradyrhizobium.
REFERENCE  1 (bases 1 to 1830)
AUTHORS    Gabel,C.
TITLES     Direct Submission
            Submitted (07-AUG-1990) Gabel C., Johns Hopkins University, Dept.
            of Biology, Charles and 34th Sts., Baltimore, MD 21218, USA
2 (bases 1 to 1830)
AUTHORS    Gabel,C. and Meier,R.J.
TITLES     Nucleotide sequence of the coxa gene encoding subunit I of
            cytochrome a3 of Bradyrhizobium japonicum
            Nucleic Acids Res. 18 (20), 6143 (1990)
JOURNAL    MEDLINE
FEATURES   91045095
            Location/Qualifiers
            1..1830
            /organism="Bradyrhizobium japonicum"
            /strain="LO"

```



```

RBS      /db_xref="taxon:375"
167..170 /note="ribosome binding site (pct.)"
178..1803 /note="cytochrome aa(3) subunit I"
CDS      /codon_start=1
         /transl_table=11
         /protein_id="CAA38216.1"
         /db_xref="GI:39506"
         /db_xref="SWISS-PROT:P31833"
         /translation="MARTSAAHGDHNDHGHENHAHPGMRKYVYSTNKHDKITVYLI
         FAVIACVGAAMSIARAELEMT PGVQIEHETIVVETSHGLIMIFEMVPMATIGF
         GMEVPELMGADPMAPRRNNISFMLEPASFGLLSTFEEBPGANGVAGMTWVP
         LSSGHPGPAVDPAIISLHLAGASSILGAINETITFNMRAPGMLHKMPLFVNSILV
         TVFLLSLPLVLAGAITMLLDNRNFTFFADGDPVLFQHLFMFGHPVYLLIL
         PEGMSIIVSTFESKRPVGYLGAMVAMVAGIGIEGVVAHMHMTVGSMSAOAYEVA
         ATMVIAPVTGVKIFESMIAVMGSGIIEFRAPMTWAGIEFLFTVGSGTVVLANAGVDR
         VIOERTYVVAHRYVLSGAVVAIFRAGTYRFPKRGITNETLAKHAFWYIEGVNL
         VEPQHFELSGMPKRYVDYPDAFAGMNLVSSVSGVLSGFEVLIPLCYIDAFAKRPV
         AADNPWAGATTLLEMTLPSPPEHOFVLPVQ"
BASE COUNT      282 a      606 c      566 g      376 t
ORIGIN
alignment_scores:
    Quality: 1829.00      Length: 545
    Ratio: 4.073      Gaps: 4
Percent Similarity: 82.385      Percent Identity: 61.835
alignment_block:
US-09-712-768-2 x BUCOXAG ..
Align seg 1/1 to: BUCOXAG from: 1 to: 1830
7 HisGlyHisAspHisSHISGLuLySLnglyPhePheThrArgTrpPhe 23
|||||.....:|
220 CACGGACATGACGAGACACGCCATCCAGCCGGATGCGGACGCTACGCTA 269
|||||.....:|
23 tSerThrAsnHisLysAspIleGlyLeuTyrLeuValAlaIleGly 40
|||||.....:|
270 TTCGACCAACACACAAGCATCGGACGATGCTACCTGCTCGCGCTG 319
|||||.....:|
40 aIValIleGlyPheIleSerValLeuPheThrValTyrMetArgLeu 56
|||||.....:|
320 TCGCGGCGCTATCGGTGCGCGATGTCGATCGGATCGCGGACGCTG 369
|||||.....:|
57 MetAspProGlyValIleTyrMetCysLeuGluGlyAlaArgLeuIle 73
|||.....:|
370 ATGATATCGGCGCGCTGACATCTTC..... 393
73 aAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMet 90
|||.....:|
394 .....:|CACGAGACGACACCTCAACGCTTCG 421
90 aIThrTyrHisGlyIleLeuMetMetPhePheValIleProAlaLeu 106
|||||.....:|
422 TGACCTCCACGCGCTGATCATGATCTTTCATGATGATGCGCGCATG 471
|||||.....:|
107 PheGlyIleGlyPheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaPro 123
|||.....:|
472 ATGCGGCGCTTCGCGACATGCTTCGCTGCGCTGCGCGCGCGCA 521
|||||.....:|
123 pMetAlaPheProArgMetAsnAsnLeuSerPheTrpLeuPheIleAla 140
|||||.....:|
522 CATGGCGCTTCGCGCGCATGAACAACATCTGCTGCTGCTGCGCGCT 571
|||||.....:|
140 IyThrAlaMetGlyValAlaSerLeuPheAlaProGlyIleAspGly 156
|||.....:|
572 CATTGCGGCGCTGCTGATGCTGCTGCGAGCGCGAGCGCGCGCC 621
|||||.....:|
157 LeuGlySerGlyValGlyTrpValLeuTyrProLeuSerThrArg... 172
|||.....:|
622 AACGGCGCTGCGCGCGCTGACCATGTACGTGCGCTGCGAGCTCGG 671

```

```

173 .GluAlaGlyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSer 189
|||.....:|
672 CCATCCCGCGCGCGCGAGTTCGATCTGCTGCTGCTGCTGCGCG 721
|||||.....:|
189 IyAlaSerSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsn 205
|||||.....:|
722 GTGCGCTGCTGATCTGCGGCGCATCACTTCATCCACGATCTTCAC 771
|||||.....:|
206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrp 222
|||||.....:|
772 ATGCGCGCGCGCGCGCATGACCCGCAAGAGATCCGCTGCTGCTGCT 821
|||||.....:|
222 rIlePheIleThrAlaTrpLeuIleLeuLeuAlaLeuProValIleAla 239
|||||.....:|
822 GATCGCGTGACCGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 871
|||||.....:|
239 IyAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhe 255
|||||.....:|
872 GTGCGATCACCATGCTGTCACCGACCGCAATTTGCGGACGACCTTCT 921
|||||.....:|
256 AsnProAlaGlyIleGlyAspProIleLeuTyrGlnHisIleLeuTrp 272
|||.....:|
922 GCGCGCGACGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTG 971
|||||.....:|
272 ePheGlyHisProGlyValTyrIleIleLeuProGlyPheGlyIle 289
|||||.....:|
972 CTTCGCGCATCCGAGGTGATCATCTGATCTGCGCGCTTCGCGCAT 1021
|||||.....:|
289 IeSerHisValIleSerThrPheSerLysLysProValPheGlyTyrLeu 305
|||||.....:|
1022 TCAGCGCATGCTGCTCCACCTTCGCGCGCAAGCCGCTGCTGCTATC 1071
|||||.....:|
306 ProMetValTyrAlaMetValAlaIleGlyValLeuGlyPheValTyr 322
|||||.....:|
1072 GCGATGCGCTACGCGCATGCTGCGGATGCGGCGCATGCGCTGCTG 1121
|||||.....:|
322 pAlaHisIleMetTyrThrValIleMetSerLeuThrGlnIleSerTyr 339
|||||.....:|
1122 GGGGACACCATGTCACGCTGCGGATGCTGCGGACGACGCGCTAT 1171
|||||.....:|
339 hMetLeuAlaThrMetValIleAlaValProThrGlyTyrLeuIle 355
|||.....:|
1172 TCGTCCGCGCGACGAGTGTGATGCGGCTCCACCGCGCTGAGATCT 1221
|||||.....:|
356 SerTrpIleAlaThrMetTrpGlyGlySerValGluPheLysSerPro 372
|||||.....:|
1222 TCGTGTATGCGCACATGTGCGGCGCTGATGATGATTCGCGCGCAT 1271
|||||.....:|
372 rLeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyValIle 389
|||.....:|
1272 GATCTGCGCGCTGCGGCTTCACTTCCTGTCACGCTGCGGCTGCG 1321
|||||.....:|
389 IyIleValLeuAlaGlnIleGlyLeuAspArgAlaTyrHisAspThr 405
|||.....:|
1322 GCGTGTGCTGCGGAACCGCGCGCTGACCGCGCTGCTGACGAGCAT 1371
|||||.....:|
406 TyrValIleAlaHisPheHisTyrValMetSerLeuGlyAlaIlePhe 422
|||||.....:|
1372 TACGTGCTGCGCGCATTCACATGCTGCTGCTGCTGCTGCTGCTG 1421
|||||.....:|
422 aIlePheAlaGlyIleTyrPheTyrMetProLysPheSerGlyAla 439
|||||.....:|
1422 GATCTGCGCGCGCTGACTACTGTTCCGCAAGATGACGCGCTACAT 1471
|||||.....:|
439 hProGluTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGly 455
|||.....:|
1472 ACNAGACGCGCTGCGGAAGCGCGCATCTGCTGCTGCTGCTGCTG 1521
|||||.....:|
456 AsnValThrPhePheProGlnHisPheLeuGlyArgGlnIleMetPro 472
|||||.....:|
1522 AACCTGCTGCTTCTCCCGACGACATTCCTGCGCTGCTGCGGATG 1571

```



```

472 gATGTYrILAsPYrProGLuAlaPheAlaLeuTrpaNlysValSerS 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1572 CCGCTACGTCGACTATCCGACGCGTCCGCGGACGTCGTCGT 1621
489 eRTYrGLyAlaPheLeuAlaPheAlaSerPheLeuPheLeuValIle 505
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1622 CGCTCGGCTTCACATTCGCGC.....TTCCGCGCTCGTATC 1659
506 PheValTYrThrLeuValAlaGlyArgGluThrArgPro..... 519
1660 TTCTCTATTCGCTGATCGATCGCTTGGAAAAAAGCTCCGCTCGCA 1709
520 .AsnProTgLYgluPheAlaAspThrLeuGLUTpThrLeuProSerP 536
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1710 CAATCCGTCGGCGCGCGACGCTGGAGTGGAGCGCTCCCTCC 1759
536 roProProAlaHisThrPheGLUTThrLeuProLys 547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1760 CGCGCGCTTCACATTCGAGTGGAGTGGCTCCCGCC 1794

seq_name: gb_ba:BJCOXA

seq_documentation_block:
LOCUS BJCOXA 2000 bp DNA BCT 07-NOV-1991
DEFINITION B.Japonicum cox a gene for subunit I of cytochrome c oxidase
(EC_number=1.9.3.1).
ACCESSION X54800
VERSION X54800.1 GI:39502
KEYWORDS cytochrome c oxidase.
SOURCE Bradyrhizobium japonicum.
ORGANISM Bradyrhizobium japonicum
Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
Bradyrhizobium group: Bradyrhizobium.
1 (bases 1 to 2000)
REFERENCE
AUTHORS Bott,M.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-1990) Bott M., Eidgenossische Technische
Hochschule, Mikrobiologisches Institut, Schmiedelbergstrasse 7,
CH-8092 Zuerich, Switzerland
2 (bases 1 to 2000)
REFERENCE
AUTHORS Bott,M., Bolliger,M. and Hennecke,H.
TITLE Genetic analysis of the cytochrome c-a3 branch of the
Bradyrhizobium japonicum respiratory chain
JOURNAL Mol. Microbiol. 4 (12), 2147-2157 (1990)
MEDLINE 91211625
FEATURES
source
1..2000
Location/Qualifiers
/organism="Bradyrhizobium japonicum"
/strain="USDA311b110spc4"
/db_xref="taxon:375"
/clone="PRJ3404"
75..1700
/gene="coxA"
75..1700
/gene="coxA"
/codon_start=1
/transl_table=11
/product="cytochrome-c oxidase"
/protein_id="CA38570.1"
/db_xref="GI:39503"
/db_xref="SWISS-PROT:P31833"
/translation="MATSAAAGHDHAKODHGHDEHAHPTGMRRVYSTNHKIDGTMYLI
FNYLAVIGAAWSIATRAELMYPGVOIFHEPTHYNYVFVSHGLIMFPMWPMIAGF
GMNFVLMIGAPPMAPRRNNNTSPMLPASFGILMSTFVEGEPGANGVGAGMTWVP
LSSGHPGPAVDAILSLHAGASSILGAINFTTTFNKRAQMTLHKRPDLVMSITV
TYELLSLPVLAGATMLTDNFGTTFADPGGDPVLFQHLTFEGHPEVYILIT
PGRGMSIQIATSTFSRKPFGVGLGMAVANAIGIGVVAHHNYTVGMSATQAFVA
ATVIAVPTGPKFESVIAIWMGSIIEFRAMIAVGFILFTVGAVTVGLVLAACGVR
VLOETYYVAHPIVSLGAVFAIFAGWYMYPKMGYMYNETLIAAHFMYEIGVNL
VFEPQHFGLISGMRRVYVDYPPAFAGMNYVSSVSIISGFGVLIIFYVIDAFKKVP
AGDNPWGAGATLTLEMTLPSPPPHPEVLPVPO"
75..1697
/mat_peptide
/gene="coxA"

```

```

CDS
/product="cytochrome-c oxidase"
1654..>2000
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CA38571.1"
/db_xref="GI:39504"
/db_xref="SPTREMBL:O99253"
/translation="MSLVITFALVGNAMPGHFHPVLATLSLCTIAGVGASGALNMA
LESDI"
1654..>2000
mat_peptide 304 a 669 c 614 g 413 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 1829.00 Length: 545
Ratio: 4.073 Gaps: 4
Percent Similarity: 82.385 Percent Identity: 61.835

alignment_block:
US-09-712-768-2 x BJCOXA ..
Align seg 1/1 to: BJCOXA from: 1 to: 2000
7 HISGLYHISASPHISGLIULSGINGLYPHEPETHRARGTRPHEME 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 CACGACATGACGACGACGACCCATCCGACCGGATGCGACGCTACGTCTA 166
23 tSerThrAsnHisLysAspIleGlyLeuLeuTYrLeuValAlaIleGly 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTCGACCAACACAAAGACATCGCGACGATGACCTGATCTTCGCGGTCA 216
40 aValGlyPheIleSerValLeuPheThrValTYrMetArgLeuGluLeu 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 TCGCGCGCTCATCGGTGCGCGGATCGATCGATCGATCGCGGACGCTG 266
57 MetAspProGLyValGlnTYrMetCysLeuGluGlyAlaArgLeuIleAl 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ATGTATCCGCGCGCTGACATCTTC..... 290
73 asPalaSerGlnThrCysThrAlaAsnGlyHisLeuTrpaNlysValMet 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 .....CACGACGACGACACCTCAACGCTTGTG 318
90 alhTrpTrpHisGlyIleLeuMetPhePheValGlyLeuProAlaLeu 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 TGACCTCCGACGCGCTGATCATGATCTTCTCATGGCTATCCGCGGATG 368
107 PheGLYGLYPheGLYAsnTYrLeuMetProLeuGlnIleGLYAlaProAs 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 ATGCGCGCTTCGCGCAATCGCTTCGCGCGCTGATGCGCGCGCGCA 418
123 pMetAlaPheProArgMetAsnAsnLeuSerPheTrpPheIleAlaIag 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 CATGCGCTTCGCGCGCAATCGCTTCGCTGCTGCTGCTGCGCGCT 468
140 IYThrAlaMetGLYValAlaSerLeuPheAlaProGLYAlaSerGlyGln 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 CTTTGGGCTGCTGCTCATGTCACCTTGTGAGGGCGACCGCGCGCG 518
173 .GluAlaGLYTYrSerMetAspLeuAlaIlePheAlaValHisLeuSerG 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 AACGCGCTCGCGCGCGCTGACCATGTACGCTCCGCTGCGAGCTCCG 568
173 .GluAlaGLYTYrSerMetAspLeuAlaIlePheAlaValHisLeuSerG 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
569 CCAATCCGCGCGCGCGCTGACCATGTGCGGATCTGCTGCTGATCTGCG 618
189 lYAlaSerSerIleMetGLYAlaIleAsnMetIleThrPheLeuAsn 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
619 GTGCTGCTGATCTGCGCGCGCATCAACTTCATACACGATCTTCAAC 668

```

206 MetatgAlaProGlyMetThrLeuHisLysValaProIleuPheSerPro 222  
669 ATGGCGGGCGGGGCAATGACCCTGCAACAATGCGCGCTTGCGTGGTTC  
222 TtLePheLleThraLrPheLleuLeuLeuAlaLeuProValLeuAlaG 239  
719 GATCGTGGTGGACCTGCTTCTGCGTGGCTGTTGGCGCTCCGCGTCCGCG  
239 LysAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePhe 255  
769 GTGGGATCCACCATGCTCTCAACGACGACCAATTTCGGACGACCTTCCTC 818  
256 AsnProAlaGlyGlyAspProIleuLeuThrGlnHisLleLeuThrPph 272  
819 CGCGCCGACGCGCGGCGGACCCCGTGGCTGTTCAGACACTGTTCTGGTT 868  
272 ePheGlyHisProGlyValTyrLleIleIleLeuProGlyPheGlyIle 289  
869 CTTTGGCCATTCGCGAGGTACATCTGATCCGATCCGCGCGCTTCGCATGA 918  
289 LeSerHisValaLserThrPheSerLysLysProValPheGlyTyrIleu 305  
919 TCAGCCAGATCGTCTCCACTTCTCCGCAAGCCCGTGTTCGGCTATCTC 968  
306 ProMetValTyrAlaMetValaLalaIleGlyValLeuGlyPheValaTyr 322  
969 GGCATGGCCATCCCATGTCGCGATCGCGCATGCGCATGCGTCTGTGGTGG 1011  
322 pAlaHisMetTyrThrValGlyMetSerLeuThrGlnGlnSerTyr 339  
1019 GGGGCAACACATGTACAAGGTGGATGTCGTCGGGACGACGCGCTAT 1066  
339 hemeLeuAlaThrMetValaLalaValProThrGlyLleLysIlePhe 355  
1069 TCGTCCGCGGACGATGCGATGCGCGGTTCCAGCGCGCTGAAGATCTTC 1116  
356 SerThrIleAlaThrMetThrProGlyGlySerValGluPheLysSerPro 372  
1119 TCGTGATCGCCACGATGTGGGGCGGCTCGATCGAATTCGCGCGCCGAT 1166  
372 LLeuThrAlaPheGlyPheMetPheLeuPheThrValGlyGlyValaTrng 389  
1169 GATCTGGCGCGTGGGCTCATCTCTCTCTCTTCACGCGGCGGTGGCGTAC 1211  
389 LylleValLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyr 405  
1219 GCGTCTGCTGGCGACGCGCGGCGTGCACGCGCTCTCCAGAGACCTTAC 1268  
406 TyrValaValaHisPheHisIstYrValaMetSerLeuGlyAlaIlePheAl 422  
1269 TACGTCGTGCGCACCTCCACTACGTGCTGTGCGTGGTGCACGTCTTGC 1318  
422 allPheAlaGlyIleTyrPheTyrMetProLysPheSerolYarGlaP 439  
1319 GATCTTCCGCGGCTGCTACTACTGTGTTCCGAAATGACGGGCTTACATGT 1368  
439 heProGluThrAlaLalaLysLeuHisPheThrPThrPheHisLleGlyAla 455  
1369 ACAAGACAGCGCTGGCGAAGCGGCACTTCTGGGTCACTTATGAGCGGTG 1418  
456 AsnValIthrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProAr 472  
1419 AACCTGGGTCTTCTCCGACACACTTCTCGCGCTGTGGGACATGCCGCG 1468  
472 gArgTyrIleAspTyrProGluValaPheAlaLeuThrPAsnLysValaSer 489  
1469 CGCGTACGTCGACTATCCCGACGCGTTCGCGCGGTGAACCTGGTCTCGT 1518  
489 eTyrGlyAlaPheLeuAlaPheAlaSerPheLeuPheAlaIleValaIle 505  
1519 CGGTGCGCTCTTACATTCCGCGC.....TTCCGCGCTCTTATC 1556  
506 PheValTyrThrIleuValaLalaGlyArgArgGluThrArgPro..... 519

```

1557 TTTCCTATATGCGTGCATGCATGCTTTGCCGAAAAAGGCGCGCGTGGCGA 1606
520 ..AsnProtrprGlyGluPheAlaAspThrLeuGluTrpThrLeuProSerp 536
1607 CAATCCGATGGGGTGGCGGACACCCCTGAGTGCACGCTGGCCTCGC 1656
536 rGpPrpAlaHisThrPheGluTrpThrProLys 547
1657 CGCGCGCCCTTCATCAGTTCCAGATGCTGCCCCGC 1691
seq_name: gb_ba:BUA242592

```

LOCUS	8121 bp	DNA	BCT	30-SEP-1999
DEFINITION	Bradyrhizobium japonicum coxB, COXA, COXE, COXF, COXG, COXC, COXD			
ACCESSION	AF242592			
VERSION	AF242592.1	GI:6006409		
KEYWORDS	assembly protein; coxA gene; coxB gene; coxC gene; coxD gene; coxE gene; coxF gene; cytochrome c oxidase subunit I; cytochrome c oxidase subunit II; cytochrome c oxidase subunit III; heme A synthase; Heme O synthase; hypothetical protein; ORF133; protoheme IX farnesyltransferase; sbd1 gene; SDR1 homolog; tldD gene; tldD homolog.			
SOURCE ORGANISM	Bradyrhizobium japonicum. Bradyrhizobium japonicum Bacteria: Proteobacteria; alpha subdivision: Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium.			
REFERENCE	1 (bases 1 to 8121)			
AUTHORS	Rossmann, R., Lotterer, H., Rossi, P. and Hennecke, H.			
TITLE	Factors involved in biogenesis of active cytochrome a3 encoded by the coxB/coxC gene cluster from Bradyrhizobium japonicum			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 8121)			
AUTHORS	Rossmann, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-MAY-1999) Rossmann R., Eidgenossische Technische			

FEATURES	Location/Qualifiers
source	1. .8121

gene	complement(1. .440) /gene="tldp"
CDS	complement(<1. .440)

```

/function= putative role in regulation of carbon storage"
/codon_start=1
/transl_table=11
/product="T1ad homolog"
/protein_id=CAB56816.1"
/db_xref=GI:6006410"
/translation="MTNPPTSLIDRANLIDRDQVRHEVARGLAGADGELFLEYSGT
ALMEDGRKQKQVYTSQGFGLRAVKDPAVQVSHSSQVSPALIRADAADAAVRCVQY
GSSSAPPHNWRGLGDNPELDAPGFEETKVKLLAIDAVLRDDP"
complement(542..1075)

```

```

/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAB56817.1"
/db_xref="GI:6006411"
/translation="MGLSRCAAPDPSSINLIGLAAIALLPSLAHQCAVRSHR
DMQIRDPDPGAQARQCALIOSVAEDRSNGLIYIYLTADQSKLMVAAPLGL
PSGLGIKEDNDVGRAGFVRLPNGVAEYVMDDKILGQILRTAKTATFIIFET
GPEPLSLINGEGEYDKI"
1523. 2362
/gene="cox8"
1523. 1612
/sig_peptide
/gene="cox8"

```



```

2593 TCGCGGCGGATCGGTCGCCGATGCGATCGGATCGGCGGACGCTG 2642
57 MetAspProGlyValGlnThrMetCysLeuGlnAlaArgLeuIleAl 73
2643 ATGATATCGGCGGTCGACATCTTC..... 2666
73 AspAlaSerGlnThrCysThrAlaAsnGlyHisLeuThrAsnValMet 90
2667 .....CACGAGACGACACCTACACAGCTTCG 2694
90 ALTHrHisGlyIleLeuMetThrPheValGlyIleProAlaLeu 106
2695 TGACCTCCACAGGCGCTGATCATGATCTTCATGATGATCGGCGCATG 2744
107 PheGlyGlyPheGlyAsnThrLeuMetProLeuGlnIleGlyAlaProAs 123
2745 ATCGCGGCGCTTCGCGCACTGCTGCGCGCTGATGATCGGCGCGCGCA 2794
123 PheAlaPheProArgMetAsnAsnLeuSerPheThrLeuPheIleAla 140
2795 CATGGCGTCCGCGCATGAACAATCGTCTGCTGCTGCTGCGCGCT 2844
140 LysThrAlaMetGlyValAlaSerLeuPheAlaProGlyAspGlyGln 156
2845 CCTTCGCGGCTGCTGCTGATGTCGACCTTCGTCGAGGCGGACGCGCGC 2894
157 LeuGlySerGlyValGlyThrValLeuThrProProLeuSerThrArg. 172
2895 AACGCGCGGCGCGCTGCGCGCATGATGATGCGCGCTGCGCGCTGCGG 2944
173 GlnValGlyThrSerMetAspPheAlaIlePheAlaValHisLeuSer 189
2945 CCATCCGCGCGCGGACATCGATTCGCGATCTGCTGCTGCTGCGCGG 2994
189 LysAlaSerSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsn 205
2995 GTGCGCTGCTGATCTCGCGCGCATCACTTCATACACGATCTTCAC 3044
206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerThr 222
3045 ATGCGCGCGCGGCGCATACCTCGACAGATGCGCGCTTCGTCGTCG 3094
222 rIlePheIleThrAlaThrLeuIleLeuLeuAlaLeuProValLeuAla 239
3095 GATCGTGTGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3144
239 LysAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhe 255
3145 GTGCGATACACATGCTGCTGACCGCGCAATTTCGCGACGACCTTC 3194
256 AsnProAlaGlyGlyGlyAspProIleLeuThrGlnHisIleLeuThr 272
3195 GCGCGCGAGCGGCGCGCGCGCTGCTTCACAGCCGCTTCGTCGTT 3244
272 PheGlyHisProGlyValThrIleIleIleLeuProGlyPheGlyIle 289
3245 CTTCGCGCATCCGCGAGTGTACATCTGATCTGCGCGCTTCGCGCAT 3294
289 IeSerHisValIleSerThrPheSerLysProValPheGlyThrLeu 305
3295 TCAGCGCAATGCTTCACCTTCGCGCGCAAGCCGCTGCTGCTGCTG 3344
306 PrometValThrAlaMetValAlaIleGlyValLeuGlyPheValIle 322
3345 GCGCATGCGCTTCGCGCATGCTGCGCGATCGCGCGCTTCGTCGTCG 3394
332 PAlaHisIleMetThrValGlyMetSerLeuThrGlnGlnSerThr 339
3395 GCGCGACACATGTACAGGTGCGCATCTGTCGCGCGACGAGCGCAT 3444
339 hemetLeuAlaThrMetValIleAlaValProThrGlyIleLysThr 355

```

```

3445 TCGTCGCGGAGACATGATGATCGCGCTTCGACCGCGCGTGAAGATCTTC 3494
356 SerThrIleAlaThrMetThrGlyGlySerValGlnPheLysSerPro 372
3495 TCGTGATCGCGCGATGAGGCGGCTGATGATGATTCGCGCGCGCAT 3544
372 tLeuThrAlaPheGlyPheMetPheLeuPheThrValGlyValIleThr 389
3545 GATTCGGCGCGGCTGATCATCTCTCTTCACGTCGCGCTGCGCGTAC 3594
389 LysIleValLeuAlaGlnAlaGlyLeuAspArgAlaThrHisAspThr 405
3595 GCGTCGCTGCGCGAAGCGCGCGCTGCGCGCTGCTGCTGCTGCTGCT 3644
406 ThrValValAlaHisPheHisThrValMetSerLeuGlyAlaIlePhe 422
3645 TACGTCGTCGCGCATTCATGATGATGATGATGATGATGATGATGAT 3694
422 AliePheAlaGlyIleThrPheThrMetProLysPheSerGlyArgAla 439
3695 GATCTTCGCGCGCTGCTGATGATGATGATGATGATGATGATGATGAT 3744
439 heProGluThrAlaAlaLysLeuHisPheThrThrPheIleGlyAla 455
3745 ACAACGAGACGCTGCGAAGCGCGCTGCTGCTGCTGCTGCTGCTG 3794
456 AsnValThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetPro 472
3795 AACCTGCTGCTTCTCCGCGACCTTCCTGCTGCTGCTGCTGCTGCT 3844
472 GATGlyrIleAspThrProGlnAlaPheAlaLeuThrPheAsnLysVal 489
3845 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3894
489 eTrpGlyAlaPheLeuAlaPheAlaSerPheLeuPheIleValIle 505
3895 CGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3932
506 PheValThrThrLeuValAlaGlyArgGlnThrArgPro..... 519
3933 TTCCTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 3982
520 AsnProThrGlyGlnPheAlaAspThrLeuGlnThrThrLeuProSer 536
3983 CAATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4032
536 roProProAlaHisThrPheGlyThrLeuProLys 547
4033 GCGCGCGCTTCATGATGATGATGATGATGATGATGATGATGATG 4067
seq_name: gb_BUJ33883
seq_documentation_block:
LOCUS BUJ33883 31495 bp DNA BCT 10-AUG-2000
DEFINITION Bradyrhizobium japonicum putative epoxide hydrolase Ephb (ephb),
putative stress-induced protein Ohr (ohr), putative transmembrane
transcriptional regulator protein TtrR (ttrR), putative
two-component system transcriptional response regulator Tcsr
(tcsr), integral inner membrane metabolite transport protein Muba
(muba), orf157, and orf235 genes, complete cds; Fo Atrase operon,
complete sequence; probable acetyltransferase Pact (pact),
mitochondrial processing peptidase-like protein Mpp (mpp),
threonine synthase Thrc (thrc), two-component system
transcriptional regulator Elmr (elmr), putative two-component
system sensor histidine kinase Elms (elms), SUR1-like protein
(shb1), and orf133 genes, complete cds; cox operon, complete
sequence; orf177, Tldd (tldd), signal peptidase Sips (sips), and
orf175 genes, complete cds; and putative glutathione synthetase
Gsmh (gsmh) gene, partial cds.
ACCESSION U33883
VERSION U33883.2 GI:8708897
KEYWORDS
SOURCE Bradyrhizobium japonicum.

```

ORGANISM Bradyrhizobium japonicum  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

REFERENCE 1  
AUTHORS Muller, P., Ahrens, K., Keller, T. and Klauke, A.  
TITLE A *tnpA* insertion within the *Bradyrhizobium japonicum* *sipS* gene, homologous to prokaryotic signal peptidases, results in extensive changes in the expression of PBM-specific nodulins of infected soybean (Glycine max) cells  
Mol. Microbiol. 18 (5), 831-840 (1995)

JOURNAL MEDLINE  
PUBMED  
825087  
2 (bases 1 to 31495)

REFERENCE 2  
AUTHORS Mueller, P.  
TITLE Extended sequencing of a DNA fragment of *B. japonicum* adjacent to the *cox* operon  
Unpublished  
3 (bases 1 to 31495)

JOURNAL MEDLINE  
PUBMED  
825087  
3 (bases 1 to 31495)

REFERENCE 3  
AUTHORS Mueller, P.  
TITLE Direct Submission  
Submitted (15-AUG-1995) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany  
4 (bases 1 to 31495)

REFERENCE 4  
AUTHORS Mueller, P.  
TITLE Direct Submission  
Submitted (26-JUN-2000) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany  
Sequence update by submitter  
On Jun 26, 2000 this sequence version replaced gi:1200339.

REMARK COMMENT  
FEATURES  
source location/Qualifiers  
1..31495  
/organism="Bradyrhizobium japonicum"  
/strain="USD1105pc4"  
/db\_xref="taxon:375"  
/clone="C27"  
complement(7..909)  
/gene="ephb"  
complement(7..909)  
/gene="ephb"  
/codon\_start=1  
/transl\_table=11  
/product="putative epoxide hydrolase Ephb"  
/protein\_id="AAF78792.1"  
/db\_xref="GI:8708901"  
/translation="MKPFTIKQIDAGVSMAMRGVPSNGVPAICCTAGPYDIAFVDVADPGKGRVYIIPYLRGTTHFLSSPTPQWRAPAMAVDIIALMDKLDIKKAVAGFDGARTADIIAALPDGAARMCRSAGYLSSQAGSAPLPSPAELOMWQFPAVEGRAGYETKYTHDFKLIMKLASPMKEDDAFFDSAAALDNKDHVAITIHNYRMLGIAKGEAKYELEKRLAAAPLISVPTMEGDANGAPHPDPSTYAKKFGSRHDYRLITGCVGNLPRAPQAPAKAVIDADCA"  
complement(1071..1490)  
/gene="Ohr"  
complement(1071..1490)  
/gene="Ohr"  
complement(1071..1490)  
/gene="Ohr"  
/transl\_table=11  
/codon\_start=1  
/product="putative stress-induced protein Ohr"  
/protein\_id="AAF78793.1"  
/db\_xref="GI:8708902"  
/translation="MTQAAKLLYTAHTHTGGGRODQMSKSDGRDLVRLSPGAGIGTNEOLFAGMSACFEGAMEIARRKRLITLPRKCAIDAEIDLLDEGAYSLARLNVSLGDLDEVARLDIIDEHQTCTPSKAVRGIDVAVALT"  
1690..4398  
/gene="ttrr"  
complement(1690..4398)  
/gene="ttrr"  
1690..4398  
/transl\_table=11  
/codon\_start=1  
/product="putative transmembrane transcriptional regulator protein Ttrr"  
/protein\_id="AAF78794.1"

misc\_feature  
/db\_xref="GI:8708903"  
/translation="MTEPAAPVAGPLSPFGPTVYIPHERLVTRDGVALLPGTKAPFLIALMSRRENVYSKMDLALVMPGAVPEANLREVALARKVLIDGKRGARYITLISGRGYCEVAPISQTDVPAERPPAGMPCPVKIPNLRQVNGDEVVAALSKDLTSHSVYTVGGGQKTAIVAAVAHADLLETFADAAHVVDLAADSDPOLVITLILMLGLAQTDDVAAHRCISOTRCCLILNCEHVAIAAPLEIAEHAPPHHILATSERLRVEGOVYRLADLAVPDAGLTAAVQYPAALQFLERATVGAQOAILDNDNAIVGILSRVDMALIELAAGREAVYGEOTFALDERILNLMOGPAPAPROKTIQATLMSYGLSGTEVLRRLAVAFAGHTIDALEVYPPDRVDRARLPADISLVAKSMVAPRRGAMMYRLDITTRAVILLETPEDDAALAGRORLIPAMGTGRDHMCNADADEAOSTFALHNVAALDWCFGANDACIGIALAAAPLFLTAMSLITCRKRSKSEALLATIDPSTRGAEEMHIOALGILMTGRKGRSGARRIEQKPCDRGGARRANQDLQCRMHITFERMGOFDAALIGYAROSITVYAEVLGDAASLGLASLGLVSLHLAGHRDALNMLEAAWQPGTERTSYHGDRHRNRAIGISLAREIMLOGRPGDARQLOTVEAQMDDPITLCTALIMVSIIDMSGELDAEESIDHPTAHATSMGPLYLSVGVGKELAIRGDAAGVETIRACLRTHDAGYEILTFPPNIALVHIGILALGOVQASRLIDALIRLVGGDHLVMPELLRMRKVLISLPERTGTGAETCLMQSLSTSGQCESLGAASG"  
2158..2217  
/gene="ttrr"  
/note="putative"  
/note="transmembrane-region site"  
complement(4540..5211)  
/gene="tcsr"  
complement(4540..5211)  
/gene="tcsr"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system transcriptional response regulator Tcsr"  
/protein\_id="AAF78795.1"  
/db\_xref="GI:8708904"  
/translation="MNAFSTHLYIADHPLEFRLDAQVAVSLTSARIDAGSFEDLTALKEQTSDDVDLIDLSMPGISGSLIYRAQYPAIPVYIVASDSATIRSDFGKLEIPIKRGVESYRDLALTKVMSGDWVPADTQSLATDPDMLRDLRVTLTPQOVRVLMMLSPGLNKQIATYELGSEATIKAHVSAIIQKLGVESTRQAVIAAARLAGQMKQGSTG"  
complement(5050..5052)  
/gene="tcsr"  
/note="putative; phosphorylation site"  
5518..7185  
/gene="mtba"  
5518..7185  
/gene="mtba"  
/codon\_start=1  
/transl\_table=11  
/evidence="not experimental"  
/product="integral inner membrane metabolite transport protein MtbA"  
/protein\_id="AAF78796.1"  
/db\_xref="GI:8708905"  
/translation="MSTMAVSOAGVGMTRDERFVLIASSLGVFEWYDFLYGLSASIIIGAOFSAVPPATRODIFALLAFAGFLVPEFGAIVGRIQDIVGRKFTLVYILINGLSTFIVGLIPPAATIGIAPIITLIALRLRQGLAGGYGAATVYVAHPNKGKGYTSFIQTRATICLPIELIVITPFRSALTEADPAMGWIPLVSLIIGISVQVIRLRNESPITQKMDGKSSKAPLVEAFENQNKVILALLGLGVMGQGVVYVITQAPLPLQSLIKVDGYTANLLIAMSLLGTGFTVFGMLSKITRKPILIGGCLIALITPPLEKMTISNANPALEKATEEGQGRGGRGRLRNVQGRRLARALRHRTLRPRVVGQVFDIGCGFRFREDGQSGSLREBQCSGSRGRRGRRLRDRRRRHEDASVHLPSAAVATIGLFTLVITVWVGPDIAMIVELPRTIRYTSMLPYHINMGFGGLPATPAIYASGDIYAGIMYVIVASTIVVIGFLFLETRDVIDKAI"  
complement(7350..7823)  
/gene="Orf157"  
complement(7350..7823)  
/gene="Orf157"  
/note="Orf157"  
/codon\_start=1  
/transl\_table=11  
/evidence="not experimental"  
/product="unknown"  
/protein\_id="AAF78797.1"  
/db\_xref="GI:8708906"  
/translation="MTSARSVTPICGAILIGASHHYLIVAAHRGDGARGLHCPDPTAFAIEARARCAVTLPIRHTSASLTIOENADPSVLDITPALSRLAPENAGMTHTDPECD

```

gene      DMPGHTKTMVTSQVPLVNGRLALGTWQALTYLLIENHRRPHREVVLQDTGANG"
complement(7853..8560)
/gene="orf235"
complement(7853..8560)
/gene="orf235"
/note="orf235"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="unknown"
/protein_id="AAF78798.1"
/db_xref="GI:8708907"
/translation="MPGRIADVPARRGCEDETSDVVHNIMAFSGAGRGDQQR
SKRAIRARRRRRLQHDADPRGAAGSTPARVLDLQERRRLGLRGARRARAR
RDRPAPDRSREAGTARAGTDVVKIDIALPKSDPKKADKSPRATODRDPARER
RSCDVVPGKGLASADRGCQPRAGRGVEDLSLSWRKRGHIRGLSHVSNALPPFA
VTRINGSNLPCALAH"
8893..9240
/gene="unc1"
/note="orf115"
8893..11493
/gene="Fo ATPase operon"
/note="constitutively expressed"
8893..9240
/gene="unc1"
/note="orf115; initial gene of the Fo ATPase operon"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="unknown"
/protein_id="AAF78799.1"
/db_xref="GI:8708908"
/translation="MELAIIDRRPNVRPNFNGSEILISGPNFPAATSRPSKPDGDR
AARASAMALGRLSSELVAGVYVAGIOMGFDRILSTSPFGFIVFLLEGVAGVYVNV

```

alignment\_scores:

Quality:	1829.00	Length:	545
Ratio:	4.073	Gaps:	4
Percent Similarity:	82.385	Percent Identity:	61.835

alignment\_block:

US-09-712-768-2 x BUJ33883/rev ..

Align seg 1/1 to reverse of: BUJ33883 from: 1 to: 31495

```

7 HISGLYHISASPHISHSGLIUSYGLNGLYPHEPHETHRATGTPHEME 23
|||||
24557 CACGGACATGACGACGACCCATCCGACCGGATGGCGGCTTCA 24508
23 tSerThrAsnHisLysAspIleGlyLeuLeuTyrLeuValAlaIleGly 40
|||||
24507 TTCGACCAACACAGACGACATCGGACGATGATCGATCTTCGCGGTCA 24458
40 aLYaIGLYPHEIleSerValIleuPheThrValTyrMetArgLeuGluLeu 56
|||||
24457 TCCCGCGGCGATCGGTGCGCGCATGTCGATCGCGATCGCGAGCTG 24408
57 MetAspProGlyValAlaGlyTyrMetCysLeuGluGlyAlaArgLeuIleAl 73
|||||
24407 ATGTATCCGGCGCTGCACGATCTTC ..... 24384
73 aAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTyrAsnValMetY 90
|||||
24383 ..... CACGAGACGACACACCTACACGCTTCG 24356
90 aLYTHYrHISGLYIleuMetMetPhePheValGlyIleProAlaLeu 106
|||||
24355 TGACCTCCACGCGCGCATGATGATCTTCTTCATGTCATGCCGCGGATG 24306
107 PHEGLYGLYPHEGLYASNTYrLeuMetProLeuGlnIleGlyAlaProAs 123
|||||
24305 ATCGGCGGCTTGCGCAACTGTGTCGTGCGCATGATGCGCGCGCCGCA 24256

```

```

123 pMetAlaPheProArgMetAsnAsnLeuSerPheThrPhePheIleAla 140
|||||
24255 CATTGGGCTTCCCGCGCATATACACATCTCTGCTTGGCTGGCCGGCTT 24206
140 IYTHAlaMetGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGln 156
|||||
24205 CCTTCGGCGCTGCTGATGTCGACCTTCTGAGGCGGACCGCGCGCC 24156
157 LeuGlySerGlyValGlyTyrPheValLeuTyrProProLeuSerThrArg 172
|||||
24155 AACGGCTCGCGCGCGCTGACGACATGACGTCGCGCTGCACGCTCCGG 24106
173 .GluAlaGlyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerG 189
|||||
24105 CCATCCCGGCGCGCGGACGATGACTTTCGCGATCTGTGCGCTGCGGCGG 24056
189 IYAlaSerSerIleMetGlyAlaIleAsnMetIleThrThrPheLeuAsn 205
|||||
24055 GTGCTCGTCGATCCTCGCGCGCATCACTTCATCACCACGATCTTCAAC 24006
206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerThrSe 222
|||||
23955 GATCCGCGTGCACCGTGTCTCTGCTGCTGCGTGGCGGCTGCTGCGCG 23906
222 rLlePheIleThrAlaTyrPheIleLeuLeuAlaLeuProValLeuAlaG 239
|||||
23905 GTGCGATCCACCATGCTGCTCACCGACCGCAATTCGCGACACACTTTC 23856
239 IYAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePhe 255
|||||
23856 ASPProAlaGlyGlyIYAspProIleLeuTyrGlnHisIleLeuTyrP 272
|||||
23855 GCGCGCGACGCGCGCGCGGACCCCTGCTGTCCAGCACTGTCTGCTT 23806
272 ePheGlyHisProGlyValIYrIleIleIleLeuProGlyPheGlyIle 289
|||||
23805 CTTCGCGCATCCCGAGGTGATCATTCGATCGTCCGCGCTTCCGACATGA 23756
289 IeSerHisValIleSerThrPheSerLysLysProValPheGlyTyrIleu 305
|||||
23755 TCAGCAGATGCTCTCCACCTTCTGCGCAAGCCCGTGTGGCTATCTTC 23706
306 ProMetValTyrAlaMetValAlaIleGlyValLeuGlyPheValIleTr 322
|||||
23705 GGCATGGCTCTACGCGATGTCGCGCATCGCGGATCGGCTTCTGCTGCTG 23656
322 PAlaHisHisMetTyrThrValIleGlyMetSerLeuThrGlnInsSerTyr 339
|||||
23656 GCGCGACCATGTTACACGAGTCCGATGTCGCGCGACGACGACGCTATTT 23606
339 hemeLeuAlaThrMetValIleAlaValProThnGlyIleLysIlePhe 355
|||||
23605 TCGTCGCGCGGACGATGATGCGGTTCCGACCGGCGGTAAATATCTTC 23556
356 SerTrpIleAlaThrMetTrpGlyIYSerValGluPheLysSerProme 372
|||||
23555 TCGTCGATCGCCACAGATGTGGGCGGCTCGATTCGATTCGCGCGCGCAT 23506
372 tLeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyValIleArg 389
|||||
23505 GATCTGGCGCGCTGCTCATCTTCCTTCCACGCTGCGTGGCGCTCACCG 23456
389 IYIleValLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyr 405
|||||
23455 GCGTCTGCTTGGCAACCGCGGCTGACGCGTGTCCAGGAAACCTTAC 23406
406 TyrValAlaAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAl 422
|||||
23405 TACGTCGTCGCGCACTTCCACTAGTGTCTGTCGCTGCTGACAGTTCGCG 23356
422 aIlePheAlaGlyIleTyrThrPheTyrMetProLysPheSerGlyValArgAla 439

```

```

|||||
23355 GATCTTCGGCGGCTGCTACTAGTGTCCGGAAGATGACGCGCTACATGT 23306
439 heptoglutrinalaialysleuHisphetRphrphetheleiala 455
23305 ACAACAGACAGCGTGGCGAAGCGCGACTTCTGGTCCACTTCACTTGGCGG 23256
456 AsnValThr.PhepheProGlnHisphLeuGlyArgGlnGlyMetProAr 472
23255 AACCTGGTGTCTTCCCGCAGCATCTCTCGGCTCTCGGCGATGCGCGG 23206
472 garGlyrIleAspTyr-ProGlnAlaPheAlaLeuTyrPasnLysValSerS 489
23205 CCCTACGTGACTATCCGACGCGCTTCCGCGGCTGGAACGTGTCGTCT 23156
489 eTyrTrpAlaPheLeuAlaPheAlaSer.PheLeuPhePheIleValIle 505
23155 CGGTGGCTCTTACATTCCTCGGC.....TTGCGGCTCTGTATC 23118
506 PheValTyrThrLeuValAlaGlyArgGluThrArgPro..... 519
23117 TTCTCTATTCGTCGATCGATCGCTTTCGAAAGAGTGGCGCTGCGCA 23068
520 .AsnProTyrGlyGluPheAlaAspThrLeuGluTyrPThrLeuProSerP 536
23067 CAATCGGTGGGTGGCGCGCGCACGCTGAGTGACGCTGCTCCCTCGC 23018
536 roProProAlaHisThrPheGluThrLeuProLys 547
23017 CGCGGCCCTTCATCGATTCTGAAGTGTGCGCCGCG 22983

seq_name: gb_pl:MTPSMGNS
seq_documentation_block:
LOCUS      MTPSMGNS      8535 bp      DNA      07-MAY-1996
DEFINITION P.subcordiformis mitochondrial atpA, nad3, cob, rps19, cox1 and
            trnX genes.
ACCESSION  Z47795.1 GI:633584
VERSION    Z47795.1
KEYWORDS   alpha subunit; apocytochrome b; atpA gene; ATPase; cob gene; cox1
            gene; cytochrome oxidase subunit I; nad3 gene; NADH dehydrogenase;
            ribosomal protein S19; rps19 gene; trnA-Lys; trnA-Pro; trnA-Tyr;
            trnA-Val.
SOURCE      Platymonas subcordiformis.
ORGANISM    Platymonas subcordiformis
REFERENCE   1 (bases 1 to 8535)
            Kessler,U.
AUTHORS     Thesis (1995) Institute for Plant Physiology, Justus Liebig
JOURNAL     University
REFERENCE   2 (bases 1 to 8535)
            Kessler,U.
AUTHORS     Direct Submission
JOURNAL     Submitted (17-JAN-1995) Kessler U., Justus Liebig University
TITLE       Institute for Plant Physiology Heinrich Bufl Ring 58-62 Giessen
            Germany 35392
REFERENCE   3 (bases 1 to 8535)
            Kessler,U. and Zetsche,K.
AUTHORS     Physical map and gene organization of the mitochondrial genome from
JOURNAL     the unicellular green alga Platymonas (Tetraselmis) subcordiformis
MEDLINE     (Prasinophyceae)
FEATURES    Plant Mol. Biol. 29 (5), 1081-1086 (1995)
            96145517
            Location/Qualifiers
                1..8535
                /organism="Platymonas subcordiformis"
                /db_xref="taxon:3161"
                /gene="atpA"
                /gene="nad3"
                /gene="cob"
                /gene="trnA"
                /function="ATP synthesis"
            CDS

```

```

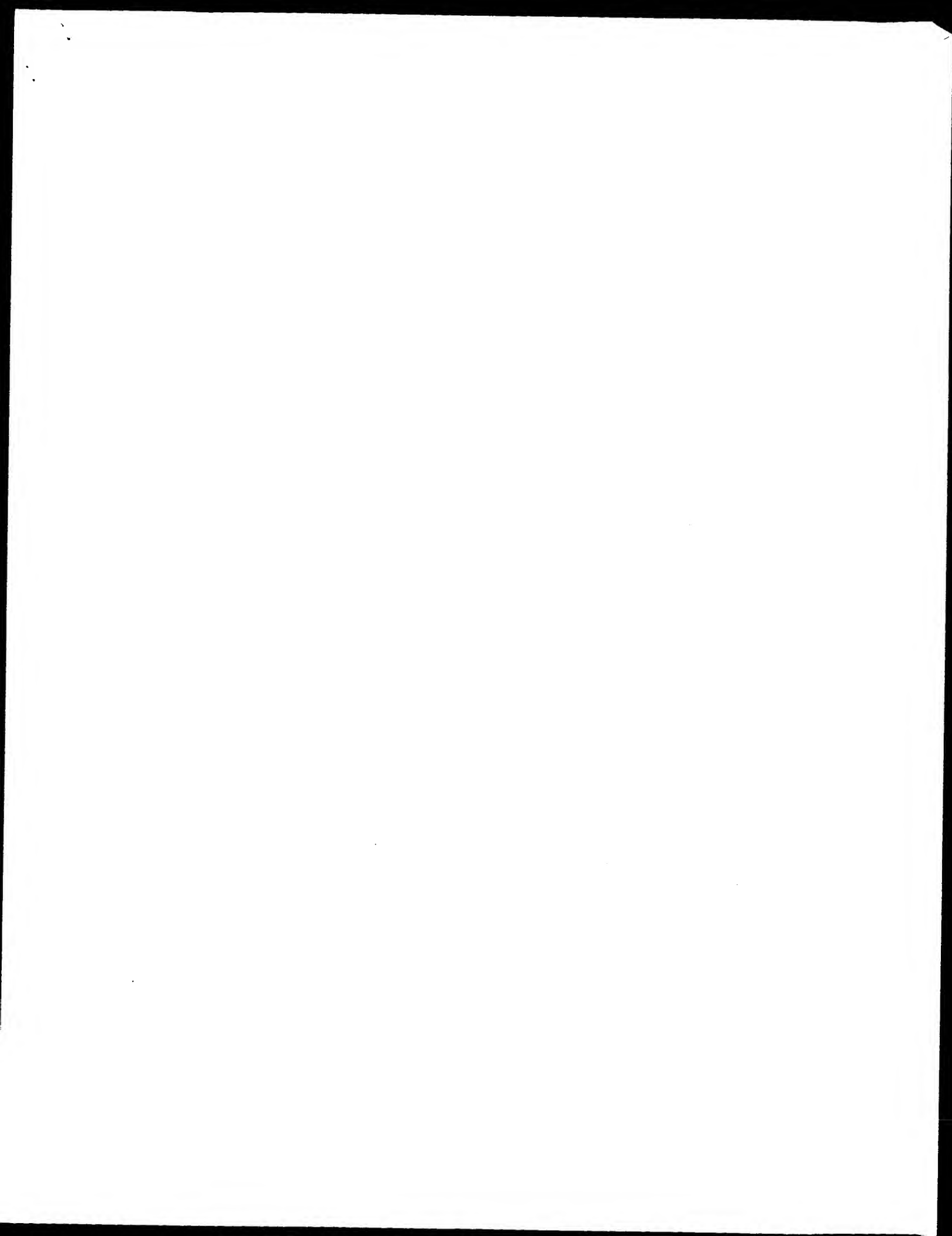
            /citation=[1]
            /codon_start=1
            /product="ATPase alpha subunit"
            /protein_id="CA87751.1"
            /db_xref="GI:633585"
            /db_xref="SPRMBL:Q36517"
            /translation="MANKRLAGELSLILRQKMEKQFNFSYEVEGKVLISGDILARY
            GLNRIDAGEWVEFQSGIKGMALNLENDNGVAVFGSDTS INEGDIIVKPTNTIVDVPV
            KGTILGVMALGNPIDGKGPLNVTQROREVAAPGILSRKSVHEPMQNGIKAVDSLVP
            IGRGQRELLIIGDRQTKTRIAIDITINOGVNEGDESKILCYVAVANGOKRSTYAO
            VKILERGALEKTSITVAATASDPAPLQFLAPYSGCMGTFPDNGMHALITIDVSKO
            SVAVRQMSLLRRPPEAREAPDQVEFLHSRLLEAKRMSDDGASLTALPVEIQAG
            DVSAYIPTNVTISITDQIFLEFELFKGIRGPAINGLSVRSYGSAQVAKAMQVAGTM
            KLELAQYREVAALAEQSDLDASTOYLNRGARLIEVLKQGLFPPMPLIENQVLLYAA
            TKGYLDKLEIKLASFEET"
            1894..2247
            /gene="nad3"
            1894..2247
            /gene="nad3"
            /function="NADH oxidation"
            /citation=[1]
            /codon_start=1
            /product="NADH dehydrogenase subunit 3"
            /protein_id="CA87750.1"
            /db_xref="GI:633586"
            /db_xref="SWISS-PROT:Q36518"
            /translation="MIEYLAVLIYFLPSLALSLITLSEIFAPQKPDPEKISAVECG
            FDPFDARCKEDIRFYVALILFLIDLEVTFLEPPAAVTLGKIGFRTGFMFLIILT
            IGRITMKRKALEME"
            2381..2454
            /gene="trnV"
            /function="protein biosynthesis"
            /product="trnA Valin (GTA)"
            2381..2454
            /gene="trnV"
            2304..3935
            /gene="cob"
            2304..3935
            /gene="cob"
            /function="electron transport"
            /citation=[1]
            /codon_start=1
            /product="Apocytochrome b"
            /protein_id="CA87751.1"
            /db_xref="GI:633587"
            /db_xref="SPRMBL:Q36519"
            /translation="MTQILTGFLAMHYTPHYHLAVSVBHIMROVEGWLIRYMAN
            GASFFIVITVIMFRGITYGTSASPREVEWICGIVITFLMLITLPAITGIVLPWGQMSFW
            GATVITISLASAIPIVGDSIVTWLMGFSVDNATLNKFSLHYLPLPIAGLSIVHIAA
            LHOYGNNPLGTGACGKVPPTPYMLKDLIGWVFISFISFVYFPYDNLGHPDNDIP
            ANPMSYPARTIVPEMYFLPYVALILRSIPKILGVLAIALVIFICLLPLINTSRISSS
            FRPLKLKLEFWLLADCFELGMIQKQVDEPDYILIGQIASVLFVYFLICIPILGIFEN
            WSVRLKDON"
            4006..4077
            /gene="trnK"
            /citation=[1]
            /function="protein biosynthesis"
            /product="trnA Lysin (GCA)"
            4006..4077
            /gene="trnK"
            4727..4978
            /gene="rps19"
            4727..4978
            /gene="rps19"
            /function="Protein biosynthesis"
            /citation=[1]
            /codon_start=1
            /product="ribosomal protein s19"
            /protein_id="CA87752.1"
            /db_xref="GI:633588"
            /db_xref="SWISS-PROT:P50893"
            /translation="MSRAIWKGPFDIDPEFRKNGSSNNKKIYSRRSVSPKFIGREV
            EIVNGHKWIKITIKEDMIGHKGEFAFTRKATILHKKTK"

```





```
431 tProLysPheSerGlyArgAlaPheProGluThrPalaIaLysLeuHisP 448
: ||||| ||| : : ||| : : : |||
7601 TGGAAATAATCTGTGCTTCATATCTGATTTTGGGACAAATTCATT 7650
||||| ||||| : : : |||
448 heTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPhe 464
||||| ||||| : : : |||
7651 TTTGGCTTTCTTTGTAGGGTTAATATTACATTTTCCCTATGCATTTT 7700
||||| ||||| : : : |||
465 LeuGlyArgGlnGlyMetProAlaArgTyrIleAspTyrProGlnAlaPh 481
||||| ||||| : : : |||
7701 TTAGGATTTCTGGAATGCCAAGCAATTCGGATTTATCTGATGCTTT 7750
||||| ||||| ||| ||||| : : : |||
481 eaIaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAla 498
||||| ||||| ||| ||||| : : : |||
7751 TGCTGATGGAATGCAGTCTGTAGTTACGGATCTTATCTTTCAGTCTTG 7800
||||| ||||| : : : |||
498 erPheIaLeuPhePheIleValIlePheValTyrThrLeuValaIaGlyArg 514
: : ||||| ||||| : : : |||
7801 GAGCTTTATTATTTTCTATGTA...GTTTACGATTCATTAAAGTGAGAA 7847
||||| ||||| : : : |||
515 ArgGluThrArgProAsnProTyr.....GlyGluPheAlaAs 527
: : : : ||||| ||| |||
7848 GAAAGATGCTCTAAATCTCTGGAAACAGTCCAGAACTTCGGCG... 7895
||||| ||||| ||| |||||
527 pThrLeuGluTrpThrLeuProSerProProAlaHisThrPheGluT 544
||||| ||||| ||||| |||||
7896 .ACATTAGAGTGGTTAACTCCGTCACCCACGCTTACCATATCTTTGAAG 7944
||||| |||||
544 hrLeuPro 546
: : : : |||
7945 AAGTTCCA 7952
```





Align seg 1/1 to: AAA91489 from: 1 to: 1674

```

1 MetAlaAspAlaAlaIleHisGlyHisAspHisHisGluGlnGlyPhe 17
1 ATGGCAGACGCCGCCATTCACGGCCATGACCACTGAGCAAGCAAGCCTT 50
17 ePheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeuLeu 34
51 CTTCACCGCGCTGCTTCATGTTCGACCAACACCAAGACATCGCTCTCTAT 100
34 YrLeuValAlaAlaGlyValAlaGlyPheIleSerValLeuPheThrVal 50
101 ACCTTGTACGCGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 150
51 TyrMetArgLeuGluLeuMetAspProGlyValAlaGlyTyrMetCysLeu 67
151 TACATGCGCCTGAGCTGATGATCGGGGTTGCTACATGATGCTTGA 200
67 uGlyAlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnGlyH 84
201 AGCGCGACGCTGTATCGCGATGCTTCGACAGCATGTACGGCAACGAGAC 250
84 LsLeuThrAsnValMetValThrTyrHisGlyIleLeuMetPhePhe 100
251 ACCTGTGGAACGTCATGTTACCTACCATGATTCGTGATGTTCTTT 300
101 ValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetProLe 117
301 GTGGGATCCCGCATTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGCT 350
117 uGlnIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeuSerP 134
351 GCAAATCGCGCTCGCGATATGCTTCCCGCTATGACAAACCTGCTGT 400
134 hePhePhePheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAla 150
401 TCTGGCTGTTCAATGCGGTACCGGATGGCGTGGCTTCCGCTGTGCCA 450
151 ProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTyrPheValLeuTyrPr 167
451 CGGGCGGTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
167 oProLeuSerThrArgGluAlaGlyTyrSerMetAspLeuAlaIlePheAla 184
501 GCCGCTGTGACCCCGCAAGCTGGGTATTCATGACCTCCGACCTTTCG 550
184 IValAlaHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIle 200
551 CGGTTCACCTGTGCGGTGCTTCGATCAGTGGCGCGCATCAACATGATC 600
201 ThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisLysValPr 217
601 ACGACCTTCTTGAAACATGCGCGCGCGCGCATGACGTGCACAAAGGCC 650
217 oLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLeuLeuAlaAla 234
651 GTTGTCTCGTGTGATCTTATCAGCGCTGCTGATCTGCTGCTGCTGCTG 700
234 euProValLeuAlaGlyAlaIlePheMetLeuLeuThrAspArgAsnPhe 250
701 TGGCGGTTCTGGCTGTCATCACTGCTGCTGACCGACCGCTTACTTTC 750
251 GlyThrThrPhePheAsnProAlaGlyGlyAspProIleLeuTyrG1 267
751 GGCACGACCTTCTTCAATCCTGCTGGCGCGGTGACCCGATCTGTACCA 800
267 nHisIleLeuThrPhePheGlyHisProGlyValTyrIleIleIleLeuP 284
801 ACACATCTGCTGTGCTTGTGGCACCGGACGATCATCATCATCTTCTGC 850
284 roGlyPheGlyIleIleSerHisValValSerThrPheSerLysLysPro 300

```

```

851 CGGCTTTGGCATCATCAGCCATGCTGTCGACCTTCTCGAAAAAGCCG 900
301 ValPheGlyTyrLeuProMetValTyrAlaMetValAlaIleGlyValLe 317
901 GTCTTCGTTACCTGCCGATGCTCTATGCAATGCTGCAATGCTGCTTCT 950
317 uGlyPheValValTrpAlaHisMetLysThrValGlyMetSerLeuT 334
951 GGGCTTTGTGCTGGCGCACCATGTACACCGTTGCTATGCTGCTGCTGA 1000
334 hArgInserTyrPheMetLeuAlaThrMetValIleAlaValProThr 350
1001 CCCAGCATCTTACTTCTATGCTGCGCACCATGTGATGCCGGTCCGAC 1050
351 GlyIleLysIlePheSerTrpIleAlaThrMetTrpGlyGlySerValG1 367
1051 GGGATTAAGATCTTCTGTCGATCGCCACGATGTGGGGGCTCGCTTGA 1100
367 uPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPheThrV 384
1101 GTTCAATCCCGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
384 AlGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAspArgAla 400
1151 TGGGTGTGTGACCGGTCATGCTGTCGCGCAAGCGGCTGACCGTGA 1200
401 TyrHisAspThrTyrTyrValValAlaHisPheHisLysTyrValMetSerLe 417
1201 TATCAGACACCATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1250
417 uGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetProLysP 434
1251 GGGTGGATCTTGGCATCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1300
434 heSerGlyArgAlaPheProGlyTrpAlaAlaLysLeuHisPheTrpThr 450
1301 TCTCGGCGCGGCTTCCCGAATGGCGTCCAAAGCTGCACTTGTGAC 1350
451 PhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuGlyAr 467
1351 TTCTTCATCGGTGCGACGTCACGTTCTTCCCGACGACCTTCCGCGAC 1400
467 gGlnGlyMetProArgArgTyrIleAspTyrProGlnAlaPheAlaLeuT 484
1401 TCGGCTATGCGCGCGCTTACATGACCTATCCGGAAGCTTTCGCTGT 1450
484 rPAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPheLeu 500
1451 GGAACAAGTCTGCTGCTATGTCGCTGCTGCTGCTGCTGCTGCTGCTG 1500
501 PhePheIleValIlePheValTyrThrLeuValAlaGlyArgArgGlyH 517
1501 TTCTTCATCGATCTTGTCTATACGCTGCTGCTGCTGCTGCTGCTGCTG 1550
517 rArgProAsnProTrpGlyGluPheAlaAspThrLeuGluTrpThrLeuP 534
1551 CCGTCCACACCGCTGGGCGCAATTCGCCGATACGTCGAATGAGACGCTGC 1600
534 roSerProProProAlaHisThrPheGluThrLeuProLysArgSerAsp 550
1601 CATCAGCGCTCCGCGCCACACGTTGGAAGCTGCGCAAGCGCTCGGAC 1650
551 TrpAspLysHisProSerHis 557
1651 TGGGACAAGCATCCCTCGCAC 1671

```

seq\_name: /sidsb/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV53011

seq\_documentation\_block:

ID AAV53011 standard: DNA: 1693 BP.

XX  
AC  
XX  
AAV53011:



```

859 GAAAAAAGAACATTGGATACATAGTGTGCTGAGCTATGATATCA 908
314 IleglyValLeuGlyPheValValIlePheValHisMetLeuThrValG 330
909 ATTGGATTCCCTAGGTTTATCGTGTGACACACCATATATTTACAGTAG 958
330 yMetSerLeuThrGlnInsertYrPheMetLeuAlaIleMetValIle 347
959 AATGACGATGACACACGACGATATTTACCTCCGCTACCATATATG 1008
347 IalaProThrGlyIleValIlePheSerTrpIleAlaIleThrMetTrp 363
1009 CTATCCCGCCCGGCTCAAAAGTATTAGTGTGATCGCAGCTCCAGCA 1058
364 GlySerValIlePheLeuSerPrometLeuTrpAlaPheGlyPheMet 380
1059 AGCAATATGAAATGATCTGCTGATGCTGATGATGATGATGATGAT 1108
380 eLeuPheThrValGlyValIleValIleValIleValIleValIleVal 397
1109 TCTTTTACCGTGTGAGTGTGCTGATGATGATGATGATGATGATGAT 1158
397 euAspArgAlaTrpHisAspTrpTrpValValAlaHisPheHisTrp 413
1159 TAGACATCGTACTACGACGACGATGATGATGATGATGATGATGATGAT 1208
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTrpPhe 430
1209 GTCTCTCATATGAGAGCTGATTTGCTCATATGAGAGCTGATCATCT 1258
430 rMetProLysPheSerGlyArgAlaPheProGlyTrpAlaIleLysLeu 447
1259 ATTTCCTCATTTCTCAGGCTACACCTAGCCAAACCTACGCCAATAATC 1308
447 IapHeTrpPhePhePheIleGlyAlaAsnValIlePhePheProGlyHis 463
1309 ATTTCATCTCATATTCATCGCGGTAATCTTCTTCTTCTTCTTCTTCT 1358
464 PheLeuGlyArgGlnGlyMetProArgTrpTrpLeuAspTrpProGly 1408
1359 TTTCTCGGCTATCGGAAATGCCCCGAGCTACTCGACCTACCCGATGC 1408
480 aPheAlaLeuTrpAsnLysValSerSerTrpGlyAlaPheLeuAlaPhe 497
1409 ATACACCATGAAACATCTATCATCTGTAGGCTCATATTTCTCTAA 1458
497 IasErPheLeuPhePheIleValIlePheValTrpThrLeuValAlaGly 513
1459 CAGCA...GTAATATTAATAATTTTCATGATTTGAGAACCTTCGCTCG 1505
514 ArgArgGlnThr.....ArgProAsnProTrpGlyLupheal 526
1506 AAGCGAAGAGCTAATAGTAGAAGAACCC.....TC 1537
526 aAspTrpLeuGlnTrpThrLeuProSerProProAlaHisThrPheG 543
1538 CATTAACCTGAGTGTATATGATATGCCCCACCTTACCAACATTTG 1587
543 Lu 543
1588 AA 1589

```

```

seq_name: /SID58/gcgdata/geneseq/NAL95.DAT: AAT05905
seq_documentation_block:
ID AAT05905 standard; DNA: 1738 BP.

```

```

AC AAT05905;
XX 29-MAY-1996 (first entry)
XX Mitochondrial cytochrome c oxidase subunit I DNA.
XX

```

```

KW Human; mitochondrial cytochrome C oxidase; COX; subunit I; subunit II;
KW subunit III; mutation: Alzheimer's disease; AD; sporadic form;
KW diabetes mellitus; IDDM; ss.
OS Homo sapiens.
EH
EH Key
EH CDS
FT Location/Qualifiers
FT 72..1613
FT /tag= a
FT /product= COX C subunit I
FT 1..71
FT /tag= b
FT 1614..1738
FT /tag= c
FT 147..149
FT /tag= d
FT /number= codon 26
FT /note= "Position of GCT>ACT, Ala>Thr mutation in AD patients"
FT 225..227
FT /tag= e
FT /number= codon 52
FT /transl_except= pos: 225..227, aa: Tyr (possible mutant)
FT /note= "Position of CAC>CCC, His>Tyr; CAC>CTC, His>Leu mutation in AD patients"
FT 267..269
FT /tag= f
FT /number= codon 66
FT /note= "Position of ATC>GTC, Ile>Val; ATC>ACC, Ile>Thr mutation in AD patients"
FT 321..323
FT /tag= g
FT /number= codon 84
FT /note= "Position of CCC>CTC, Pro>Leu mutation in AD patients"
FT 333..335
FT /tag= h
FT /number= codon 88
FT /note= "Position of GGT>GAT, Gly>Asp mutation in AD patients"
FT 378..380
FT /tag= i
FT /number= codon 103
FT /note= "Position of TGA>CGA, Trp>Arg mutation in AD patients"
FT 396..398
FT /tag= j
FT /number= codon 109
FT /note= "Position of CTC>CCC, Leu>Pro mutation in AD patients"
FT 402..404
FT /tag= k
FT /number= codon 111
FT /note= "Position of CTC>CCC, Leu>Pro mutation in AD patients"
FT 477..479
FT /tag= l
FT /number= codon 136
FT /note= "Position of TAC>CAC, Tyr>His mutation in AD patients"
FT 534..536
FT /tag= m
FT /number= codon 155
FT /note= "Position of GTC>ATC, Val>Ile; GTC>GCC, Val>Ala mutation in AD patients, position of mutation hotspot"
FT 570..572
FT /tag= n
FT /number= codon 167
FT /note= "Position of ACA>GCA, Thr>Ala mutation in AD patients, position of mutation hotspot"
FT 579..581
FT /tag= o

```



```

83  YHISLeuTripanValMetValThrTYRHisGlyIleLeuMetMetPhePhe 100
    |||.....|
224 CCACATCTACACGTTATCTGTCACAGCCCATGCTATTTGTAATATATCTTCT 273
    |||.....|
100 heValGlyIleProAlaLeuPheGlyIlePheGlyAsnTYRLeuMetPro 116
    |||.....|
274 TCATAGTAATACCATATATCGAGAGCTTTGSCAACACTGACATGTTCC 323
    |||.....|
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeuSe 133
    |||.....|
324 CTATATATCGGTGCCCCGATATGGCTTTCCCGCATTAACAACATATAG 373
    |||.....|
133 rPheTrPhePheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheA 150
    |||.....|
374 CTCTGACTTACTTACCTCCCTCTCTCTACCTGCTGCTGCTGCTATAG 423
    |||.....|
150 IaProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTPPValLeuTYR 166
    |||.....|
424 TG.....GAGCGGAGACAGAACAGGTTGACACAGTCTAC 458
    |||.....|
167 ProProLeuSerThrArg.....GluAlaGlyTYRserMetAspLeuAl 181
    |||.....|
459 CTTCCCTTACAGGAGACTACTCTCCACCTGGAGCCTCCGTAGACCTAAC 508
    |||.....|
181 aIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleA 198
    |||.....|
509 CATCTCTCTCTACCTACAGAGTGTCTCTATCTTAGGGGCAATCA 558
    |||.....|
198 smetIleThrTrpPheLeuAspMetArgAlaProGlyMetThrLeuHis 214
    |||.....|
559 ATTTCATCACACAAATATATCAATATAAACCCCTGCCATTAACCAATAC 608
    |||.....|
215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIle 231
    |||.....|
609 CAACAGCCCTCTTCGTCGTGATCCGCTAATACACAGAGCTCTACTTCT 658
    |||.....|
231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspA 248
    |||.....|
659 CTAATCTCTCCAGTCTGATGCTGCTGACACTACTACTACTAACAGACC 708
    |||.....|
248 rGAsnPheGlyThrTrpPheAspProAlaGlyGlyAspProIle 264
    |||.....|
709 GCACCTCAACACACCTTCTTGACCCCGGAGAGAGAGACCCCATTT 758
    |||.....|
265 LeuTYRGlnHisIleLeuTrpPhePheGlyHisProGluValTYRIleI 281
    |||.....|
759 CTATACCAACACCTATTTGATTTTGGTCAACCTGAGATTATATCT 808
    |||.....|
281 eIleLeuProGlyPheGlyIleIleSerHisValAlaSerThrPheSer. 297
    |||.....|
809 TATCTTACCAAGGCTTGGATATATCTCCATATTTACTTACTACTCG 858
    |||.....|
298 ..LysLysProValPheGlyTYRLeuProMetValTYRAlaMetValAla 313
    |||.....|
859 GAAAAAAGAACCATTTGGATACATAGTATGCTGACCTTATGATATCA 908
    |||.....|
314 ILeGlyValLeuGlyPheValValTYRAlaHisIleMetTYRThrValG 330
    |||.....|
909 ATTGGATTCCTAGGGTTATCTGTGACACACCATATATTACAGTAGG 958
    |||.....|
330 yMetSerLeuThrGlnIleSerTYRPhMetLeuAlaThrMetValIleA 347
    |||.....|
959 AATTAGACGTAGACACACAGCATATTTTCACTCCCTACCAATAACATTC 1008
    |||.....|
347 IaValProThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGly 363
    |||.....|
1009 CTATCCCCCAGCGGTCAAGATATTAGTGATGACGCCACACTCCACGA 1058
    |||.....|
364 GlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetPh 380
    |||.....|
1059 AGCAATATGAAATGATGCTGCAAGTCTGAGCCCTGAGATTCATCTT 1108
    |||.....|
380 eLeuPheThrValGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyL 397

```

```

1109 TCTTTTACCGGTAGTGCGCTGACTGGCATTTGATAGCAACATCATAC 1158
    |||.....|
397 euAspArgAlaTYRHisAspTrpTYRValValAlaHisIleHisTYR 413
    |||.....|
1159 TAGACATCTGACTACACGACGACGACTAGCTGTAGCCCATTCACATYT 1208
    |||.....|
414 ValMetSerLeuGlyValAlaIlePheAlaIlePheAlaGlyIleTYRPh 430
    |||.....|
1209 GTTCATATCAATAGAGCTATTTTGGCATCATAGAGGCTTCATTCACGT 1258
    |||.....|
430 rMetProLysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuH 447
    |||.....|
1259 ATTTCCTCATTTCTCAGGCTACACCTTACACCAACCAACCTACGCAAAATCC 1308
    |||.....|
447 IaPheTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
    |||.....|
1309 ATTTCATCATATTTACCGCGTAATCTTACTTCTTCCACACATAC 1358
    |||.....|
464 PheLeuGlyArgGlnGlyMetProArgArgTYRIleAspTYRProGluAl 480
    |||.....|
1359 TTTCTCGGCTATCCGGAATGCCCGACGTTACTGCACTACCCGATGC 1408
    |||.....|
480 aPheAlaLeuTrpAsnLysValSerSerTYRGLYAlaPheLeuAlaPheA 497
    |||.....|
1409 ATACACACATGAACATCTCATCTGTAGGCTCATATCTCATATTTCTTAA 1458
    |||.....|
497 IaSerPheLeuPheIleValIlePheValTYRThrLeuValAlaGly 513
    |||.....|
1459 CAGCA...GTATATTATTAATTTTCATGATTTGAGAGCCTGCTTGC 1505
    |||.....|
514 ArgArgGluThr.....ArgProAsnProTrpGlyGluPheAl 526
    |||.....|
1506 AAGCGAATAAGTCTTAATATGTAAGAAACCC.....TC 1537
    |||.....|
526 aAspThrLeuGlnTrpThrLeuProSerProProProAlaHisThrPheG 543
    |||.....|
1538 CATAACTGTGAGTGACATATGATGATGCCCCACCCCTACACACATTCG 1587
    |||.....|
543 Lu 543
    |||.....|
1588 AA 1589
    |||.....|
seq_name: /SID8/gcgdata/geneseq/geneseqn/NA199.DAT.AA211492
seq_documentation_block:
ID AA211492 standard; DNA; 16569 BP.
XX
AC AA211492;
XX
DT 02-NOV-1999 (first entry)
XX
DE Mutated human mitochondrial DNA sequence.
XX
KW Mutation; human; mitochondrial gene; abnormality detection; diagnosis;
XX diabetes mellitus; ds.
XX Homo sapiens.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT mutation 3423 /tag= a
FT /note= "native G is mutated to T"
FT mutation 3426 /tag= b
FT /note= "native A is mutated to G"
XX
PN JP11221077-A.
XX
PD 17-AUG-1999.
XX
PF 04-FEB-1998; 960P-0023196.

```







```

248 rgAsnphEgLyThrThrPhePheAsnProAlaGlyGlyAspProIle 264
|||||.....
6541 GCACCTCAACACACACCTCTTCGACCCCGGAGGAGGAGACCCCAT 6590
265 LeuTyrGlnHisIleLeuTrpPhePheGlyHisProGluValTyrIle1 281
|||||.....
6591 CTATACCAACACCTATTCGATTTTGGTCCACCTCAAGTTATATTC 6640
281 eileLeuProGlyPheGlyIleIleSerHisValValSerThrPheSer. 297
|||||.....
6641 TATCCTACGAGGCTCGAATATTCCTCATATTTGATCACTACTCG 6690
298 ..LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
|||||.....
6691 GAAAAAAGAACATTGGATACATAGATAGTGTGACCTGACCTATGATCA 6740
314 IleGlyValLeuGlyPheValValTyrAlaHisIleMetTyrThrValG1 330
|||||.....
6741 ATTGCTTCCTAGGTTATCGTGTGAGCACACCATATATTTACAGTAG 6790
330 yMeSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleA 347
|||||.....
6791 AATAGACGTAGACACACACATATTCACCTCGCTACCATATCATCG 6840
347 laValProThrGlyIleLysIlePheSerTyrPleAlaThrMetTyrGly 363
|||||.....
6841 CTATCCCAACCGCGCTCAAGATTTAGCTGACGCGACACTCCACGGA 6890
364 GlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetPh 380
|||||.....
6891 AGCAATATGAATGATCTCTGACAGTCTGACGCTGACGCTGACGCTG 6940
380 eleuPheThrValGlyGlyValThrGlyIleValLeuValGlnAlaGlyL 397
|||||.....
6941 TCTTTACCGTAGTGCGCTGACGCTGACGCTGATTTAGCAACATCATC 6990
397 euAspArgAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyr 413
|||||.....
6991 TAGACATCGACTACAGACACGCTACGCTGACGCTGACGCTGACGCT 7040
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTy 430
|||||.....
7041 GTCCATCATATAGAGCTGTATTTGCTCATATAGAGGCTTCATTCACG 7090
430 rMeCProLysPheSerGlyArgAlaPheProGluTrpAlaIleLysLeuH 447
|||||.....
7091 ATTTCCCTATTCAGCTACACCTAGACCAAACTGACGCAAAATCC 7140
447 lAsPheTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
|||||.....
7141 ATTTCAGTTCATATTCATCGGCGTAATCTAAGCTTTCTTCCCAACAC 7190
464 PheLeuGlyArgGlnGlyMetProArgArgTyrIleAspTyrProGluAl 480
|||||.....
7191 TTTCTGCGCTATCCGAATGCCCGACGCTGACGCTGACGCTGACGCT 7240
480 apheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheA 497
|||||.....
7241 ATACACCATATGAAACATCTATCATCTGAGGCTCATTCATTTCTGTA 7290
497 lAsPheLeuPhePheAlaValIlePheValTyrThrLeuValAlaGly 513
|||||.....
7291 CAGCA...GTAATATTAATATTTTCATGATTTGAGAACCTTCGCTCG 7337
514 ArgArgGluThr.....ArgProAsnProTyrGlyGluPheAl 526
|||||.....
7338 AAGCGAAGAGCTCAATATAGAGAACCC.....TC 7369
526 aAspThrLeuGlnTyrThrLeuProSerProProAlaHisThrPheG 543
|||||.....
7370 CATTAACCTGAGTGCATATATGATGATGCCCGACCTTACACACATTCG 7419
543 Lu 543

```

```

11
7420 AA 7421
seq_name: /STDSE/gcgdata/geneseq/geneseq/NM1999.DAT:AA211491
seq_documentation_block:
ID AA211491 standard; DNA; 16569 BP.
XX
AC AA211491;
XX
DT 02-NOV-1999 (first entry)
XX
DE Mutated human mitochondrial DNA sequence.
XX
KW Mutation; human; mitochondrial gene; abnormality detection; diagnosis;
diabetes mellitus; ds.
XX
OS Homo sapiens.
XX
SY Synthetic.
XX
FH Key 3426 location/Qualifiers
FT mutation /*tag= a
FT FT /note= "native A is mutated to G"
XX
PN JP11221077-A.
XX
PD 17-ANG-1999.
XX
PF 04-FEB-1998; 98JP-0023196.
XX
PR 04-FEB-1998; 98JP-0023196.
XX
PA (SAKA ) OTSUKA PHARM CO LTD.
XX
DR WPI; 1999-520716/44.
XX
PT Mutated human mitochondrial gene - useful for diagnosing diabetes
PT mellitus
XX
PS Disclosure; Page 8-14; 27p; Japanese.
XX
CC The invention relates to a DNA containing at least Base No. 3426 site of
CC a mutated human mitochondrial gene. The DNA contains a mutation at base
CC 3426 where adenine is replaced by guanine or the base is deleted. The
CC maintains identification specificity for the detection of abnormality in human
CC invention provides a method for the detection of abnormality in human
CC mitochondrial gene in which the replacement of the base from A to G at
CC the site of Base No. 3426 in the human mitochondrial gene of a patient is
CC detected. A reagent kit containing a primer that can be used for the
CC synthesis of the mutated human mitochondrial DNA can be used for the
CC diagnosis of diabetes mellitus. The present sequence represents a
CC mutated human mitochondrial DNA.
XX
SQ Sequence 16569 BP; 5122 A; 5175 C; 2178 G; 4094 T; 0 other;
XX
alignment_scores:
Quality: 1328.00 Length: 534
Ratio: 3.192 Gaps: 7
Percent Similarity: 77.903 Percent Identity: 50.187
alignment_block:
US-09-712-768-2 x AA211491 ..
Align seg 1/1 to: AA211491 from: 1 to: 16569
17 phePheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeu 33
||| .....
5907 TTGGCGAGCGTTGACTATCTCTACAAACCAACACATGACACT 5956
33 uTyrLeuValAlaAlaGlyValGlyPheIleSerValLeuPheThr 50
|||||.....

```

5957 ATACCTATTATTGCGCATGAGCTGAGTCCATGAGCAGCTCTAACCC 6006  
 50 a1TyMeTArGLeUGlUleuMeTAsPrOGlYValGlnTyMeTcysLeu 66  
 6007 TCCTATTTCGAGCGGAGCTGGCCAGCCAGGC..... 6038  
 67 GLUGlYAlArGleuIlleAlAsPrAsSerGlnThrcysThAlAsnGl 83  
 6039 .....AACCTTCAGGTAAACGA 6055  
 83 yHlSleUTrPAsnValMeTValThrTyHlSGLYlLeuMeTcysLeu 100  
 6056 CCACATCTCAACAGCTTATGCTACAGCCCATGTCATTTGTAATATCTTCT 6105  
 100 heValGlylLePrOAlALeuPheGlylPheGlyAsnTyrlLeuMeTPro 116  
 6106 TCATGATATACCCATCATATCGAGGCTTTGSCACATGACTAGTTCCT 6155  
 117 LeuGlnlLeGlyAlAPrOAsPMeTAlAPhePrOArGMeTAsnLeuSe 133  
 6156 CTAATTAATCGGCGCCGATATGCGCTTTCGCCATTAACACATTAAG 6205  
 133 PheTrPLeuPheIlleAlaGlyThrAlaMeTGLYValAlAsSerLeuPheA 150  
 6206 CTTCTGACTCTTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6255  
 150 lAPrOGlYGLYAsPGLYGLNLeuGLYSerGLYValGLYTrValLeuTy 166  
 6256 TG.....GAGCGGAGGAGGAGAACAGTTGAACACTCTTAC 6290  
 167 ProProLeuSerThrArg.....GluAlaGLYTrSerMeTAsPLeuAl 181  
 6291 CTTCCCTTACAGGAGACTTCCACCTCGAGGCTCCGTAAGCTTAC 6340  
 181 a1lPheAlaValAlAsLeuSerGLYAlAsSerSerlLeuMeTGLYAlAla 198  
 6341 CACTCTCTCTTACACCTAGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 6390  
 198 smetlIethrThrPheLeuSmetAlaGlyAlAPrOGlYMeTThLeuHlS 214  
 6391 ATTTCATCTCAACATTAATCAATTAATAAACCCCTGCCATTAACCAATAC 6440  
 215 lYsValProLeuPheSerTrPserlIethrPheIlleThrAlaTrPLeuIlle 231  
 6441 CAACGCCCT 6490  
 231 lLeuAlaLeuProValleuAlaGLYAlAlleThrMeTLeuThAsP 248  
 6491 CTTATCTCTCCAGTCTCTAGTCTGTGCTATCTATCTATCTATCTATCTATCT 6540  
 248 rGAsnPheGLYThrThrPheAsnProAlaGLYGLYlYAsPProIlle 264  
 6541 GCAACCTCAACACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6590  
 265 LeuTyrlGlnHlSIlleuTrPhePheGlyHlSPrOGlYValTyrlAla 281  
 6591 CTTATACCAACACTTATCTGATTTTTCGCTACCTGAAAGTTTATATCT 6640  
 281 e1lLeuPrOGlYPhGlylLeIlleSerHlSValAlAsSerThrPheSer 297  
 6641 TATCTTACACAGGCTCGAATAATCTCCCATTTGTATCTTACTAGTCTCG 6690  
 298 ..LyslYsProValPheGLYTrLeuProMeTValTyrlAlaMeTValAla 313  
 6691 GAAAAAACAACCATTTGATATACATAGTATGCTAGCATATATATCA 6740  
 314 lLeGlyValleuGLYPhValValTyrlAlaHlSMeTcysLeuTyrlValG 330  
 6741 ATTTGGCTTCTAGGCTTATCTGTGAGCAGACCATATATTCTACAGTACG 6790  
 330 yMeTserLeuThrGlnGlnSerTyrlPheMeTLeuAlaThrMeTValIlleA 347  
 6791 AATAGACTAGACACAGAGCATATTTCACCTCGCTACCATTAATCATCG 6840

347 lAlaVlProThrGlylLeuSylLePheSerTrPleAlaThrMeTTrPGLY 363  
 6841 CTATCCACCGCGGCTGCAAAAGTATTATGCTGACTGCCACACTCCACGGA 6890  
 364 GlySerValGlUphelySerPromleuTrPAlaPheGLYPhMeTcP 380  
 6891 AGCAATATGAAGATGCTGCTGAGTCTGAGCCCTGAGATCATCTT 6940  
 380 eLeuPheThValGLYGLYValThrGLYlLeValleuAlaGLNAlaGLY 397  
 6941 TCTTTTACCGCTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6990  
 397 euAsPrAlaGLYHlSAsPrThrTyTrValValAlaHlSPhHlS 413  
 6991 TAGACATCGTACTACAGCAGACGCTAGCTAGCTAGCTAGCTAGCTAGCT 7040  
 414 ValMeTserLeuGLYAlAllePheAlAllePheAlaGLYleryThPhe 430  
 7041 GTCTATCATATGAGAGCTGATTTTTCATCATATGAGAGCTTCACTTCA 7090  
 430 rMeTProLysPheSerGLYHlGAlaPhePrOGlYTrPAlAlaLysLeu 447  
 7091 ATTTCCTATTTCTAGGCTACACCTTACACCAACCTAGCCAAATCC 7140  
 447 lAsPheTrPThrPhePheIlleGLYAlAsnValThrPhePrOGlNlS 463  
 7141 ATTTCTATCATATATTCATGCGGCTAATATCTAATCTTCTTCCACACAC 7190  
 464 PheLeuGLYArgGlnGLYMeTProAlaGLYTrlLeAsPTrPrOGlAl 480  
 7191 TTTCTGCGCTATCCGGAATGCCCGACGCTTACTGAGTACCCGATCC 7240  
 480 aPheAlLeuTrPAsnLysValSerSerTyrlGLYAlaPheLeuAlaPheA 497  
 7241 ATACACCATGAACATCTATCATCTAGCTAGCTATCTATCTCTTCA 7290  
 497 lAserPheLeuPhePheIlleVallePheValTyrlThLeuValAlaGLY 513  
 7291 CAGCA...GTAAATTAATTAATTTATCATGATTTGAGAGCTTGGCTTGG 7337  
 514 ArgAlaGLYThr.....ArgProAsnProTrPGLYGlUphAl 536  
 7338 AAGCGAAAGTCTTAATAGTAGAAGAACCC.....TC 7369  
 526 aAsPrThLeuGLYTrThrThrLeuProSerProProAlaHlSThrPheG 543  
 7370 CATTAACCTGGAGTACTATATGATGCCCCACCTTACACATCATG 7419  
 543 lu 543  
 7420 AA 7421

seq\_name: /SIDSB/gcgdata/geneseq/geneseqn/NA2000.DAT:AA257203  
 seq\_documentation\_block:  
 ID AA257203 standard; DNA; 16569 BP.  
 AC AA257203;  
 XX  
 XX  
 DT 30-MAR-2000 (first entry)  
 DE Human mitochondrial DNA sequence SEQ ID NO:2.  
 XX  
 XX  
 KW Human; mitochondrial DNA; extramitochondrial DNA; mtDNA; exmtDNA;  
 KW diagnosis; quantification; detection; dysonomia; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease; schizophrenia; stroke;  
 KW non-insulin dependent diabetes mellitus; mitochondrial encephalopathy;  
 KW lactic acidosis; myoclonic epilepsy ragged red fibre syndrome;  
 KW Leber's hereditary optic neuropathy; ds.  
 KW  
 OS Homo sapiens.  
 XX

[illegible]

```

414 ValMetSerLeuGluValAlaIlePheAlaIlePheAlaGlyIleTyrPheTy 430
7041 GTCTTATCATATAGACGCTGATTTGCCATCATATAGAGCGCTTATTCAC 7090
430 MetProLysPheSerGlyArgAlaIleProGluTrrPalaIaLysLeuH 447
7091 ATTTCCCTCATTTCTCCAGGCTACACCTTAGACCAAACTACGCAAAATTC 7140
447 IsPheTrrPthrPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
7141 ATTTCACTCATCATATTCATCGCGGTAAATCAATCTTCTCCCAACAAC 7190
464 PheLeuGlyArgGlnGlyMetProArgArgTyrIleAspTyrProGlnAl 480
7191 TTTTCGTGCGCATCCGCGAATGCCCGCACCTTACTCTGGAGCTACCCGATGC 7240
480 apheAlaIleuTrrpsnLysValSerSerTyrGlyAlaPheLeuAlaPhea 497
7241 ATACACCAACATGAACATCTCATCATCTGATGAGCTCATTTCTGCTAA 7290
497 IsSerPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
7291 CAGCA...GTAAATATTAATATTTTCATGATTTGAGAACCTTCGCTTCG 7337
514 ArgArgGluThr.....ArgProAsnProTrrGlyLysLeuPheAl 526
7338 AAGCGAAAGCTCTAATAGTGAAGAAGACCC.....TC 7369
526 AlaPrrrLeuGluTrrPrrrLeuProSerProPropAlaIleHisrPheG 543
7370 CATTAACCTGGAGTGACTATATGATGATGCCCGCCACCTTACCAACATTTG 7419
543 Lu 543
7420 AA 7421

seq_name: /SID58/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV43718
seq_documentation_block:
ID AAV43718 standard; cDNA to mRNA; 1539 BP.
XX
AC AAV43718;
XX
XX 16-NOV-1998 (first entry)
XX
DE Cancer associated gene fragment 3.
XX
KW ds: cancer; PCR; Northern blotting; ribonuclease protection assay
XX
CS diagnosis; metastatic cancer.
XX
SS Synthetic.
XX
XX MO9837187-A1.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98MO-JP0667.
XX
PR 21-FEB-1997; 97JP-0052508.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Asada K, Hino F, Kato I, Mukai H, Yoshikawa Y;
XX
DR WPI, 1998-467552/40.
XX
PT Detection of cancer cells in tissue samples - by changes in mRNA
PT expression compared to normal tissue of specific cancer-associated
XX gene sequences
XX
PS Claim 1; Page 43-44; 92pp; Japanese.
XX

```

The cancer associated gene fragments AAV43716-V43731 can be used to detect cancer cells in tissue samples or biological fluids. They are detected by monitoring the change in mRNA expression as compared to normal tissue of one or more cancer-associated genes whose cDNA stringently hybridises to the nucleic acid fragments. The change in expression may be an increase or a decrease acid fragments. The mRNA expression may be determined by PCR, Northern blotting or ribonuclease protection assay, or by determining the change in CC of protein encoded by the gene(s) as compared to normal tissue. CC example by using a labelled antibody recognising the protein. Detection of cancer cells in tissues other than the primary tumour site.

Sequence 1539 BP; 417 A; 463 C; 249 G; 410 T; 0 other;

alignment\_scores:  
Quality: 1325.00 Length: 534  
Ratio: 3.193 Gaps: 7  
Percent Similarity: 77.715 Percent Identity: 50.187

alignment\_block:  
US-09-712-768-2 x AAV43718 ..

Align seg 1/1 to: AAV43718 from: 1 to: 1539

```
17 phephetrArGTrPheMetserThrsnHlsLysAspIleGlyLeu 33
    ||| .....|||
4 TTGGCGAGCGTTGACTATCTGTCAAAACCAACAACATGTGAACACT 53
33 uTytleuValAlalagLyvalAgLyPhelIseserValLeuphetrv 50
54 AFACCTATTATTCGGGCATGAGCGAGNCCNAGGCCACAGCTTAAACC 103
50 aTYrMeTArgLeuGIuleuMelasprroGlyValAlIntYrMetCysLeu 66
   ::::||| ||||| ::|||
104 TCCTTATTCCAGCCGAGCTGGGCCAGCCAGCC..... 135
67 GluGlYAlAArgLeuLIleAlaSPalaSerGINrCYsrPrAlaAnGI 83
136 .....|||::|::|::|
83 yHLSLeUTPasnValMeTValThrTYrHisGLyIleuMeKhePheP 100
153 CGCACATCTAACACGTTATTCGTACAGCCCATCATTTGTATATCTCT 202
100 heValIGLYIeProAlaleuPheGLyGLYPheGLyasnTYrLeuMeTPro 116
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|
203 TCATGATTAATCCCATTATTCGAGGGCTTTGGCAACTACTAGTTCC 252
117 LeuGInIIegLYAlAProAspMeLaIaehPeRofrgMeLsnInsLeuSe 133
   ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|
253 CTAAATAACGGTGCCCCGATATGGCGTTCCGCCCATTAACACATTAAG 302
133 rPhetTriPeuPheLIleagLYThrAlaMeGLyValAlaSerLeuPhea 150
   ||||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|
303 CTCTGACTCTTACCTCCCTCTCTCTACCTCCGCTCGCATATGTCTAAG 352
150 lApProGLyGLyaspGLyInLeuGLyserGLyValGLYrPvaLeuTYr 166
353 TA.....GAGCCGAGCAAGAAGACAGGTGAACAGGTCCTAC 387
167 ProProLeuSerThrArg.....GUAlaGLYrYSerMeTAspLeuAl 181
388 CTCCTCCCTTAGAGGACACTACTCCACCCCGAGCCCTCGTAGACCTAAC 437
181 alIepheaLaValHISLeuSERGLYlaSerSerILleMeTcYlAlaIEa 198
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
438 CATCTCTCTTACACTACAGAGTGTCTCTCATCTTAGGCGCATCA 487
198 smMetLIethrThrPheLeuAsmeTargAlaLAProgLYmetrPrHeuHIS 214
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
488 ATTTCATCAACAACATTTATCAATATAAACCCTCCGCATTAACCCATAG 537
```





```

75 TTCGCCGACCGTTCATCTCTACAAACCAAGACATTGGAACACT 124
33 uTYrLeuValAlaIaGlyValAlaGlyPheIleSerValLeuPheThr 50
125 ATACCTATTATTGGCGCATGAGCTGAGTCTTAGCCACACCTTAGACC 174
50 aLYrMetAlaGlyLeuGluLeuMetAspProGlyValAlaGlnTYrMetCysLeu 66
175 TCCTTATTGAGCGGAGCTGGCGCAGCAGCC..... 206
67 GluGlyAlaArgLeuIleAlaSpaIleSerGlnTYrCysThrAlaSnG 83
207 ..... 223
83 yHISLeuTrpAsnValMetValIleThrTYrHISGlyIleLeuMetMetPheP 100
224 CCACATCTACACAGCTTATGTCACAGCCCATGATTGTAAATAACTCT 273
100 heValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTYrLeuMetPro 116
274 TCATAGTATATCCCATCATATATGAGAGGCTTGGCAACGATAGTATGCC 323
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnIleLeu 133
324 CTATTAATCGGTGCCCCGATATGCGCTTCCCCGATTAACACATAG 373
133 rPheTYrLeuPheIleAlaGlyTYrAlaMetGlyValAlaSerLeuPheA 150
374 CTCTGATCTCTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATAG 423
150 laProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTYrValLeuTYr 166
424 TG..... 458
167 ProProLeuSerThrArg.....GluAlaGlyTYrSerMetAspLeuAl 181
459 CCGCCCTTACAGGAGGACTACTCCACCCCTGACCTCCGTAACCTAAC 508
181 aIlePheAlaValAlaHisLeuSerGlyAlaSerSerIleMetGlyAlaIleA 198
509 CATCTTCTCTTACACCTAGAGGAGTCTCTCTATCTTAGGGGCATCA 558
198 sMeIleIleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHis 214
559 ATTTCATACCAACAATTATCATATTAACCCCTGCCATTAACCCAAATAC 608
215 LysValAlaProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIle 231
609 CAACGCCCTCTCTGCTGATCCGCTTAATCACAGCAGCTCTACTTCT 658
231 uLeuAlaLeuProValIleuAlaGlyAlaIleThrMetLeuLeuThrAspA 248
659 CCTATCTCTCCAGTCTAGCTGCTGATCATCTACTACTAACACAGC 708
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 264
709 GCACCTCAACACACACTTCTTCCAGCCCGCGGAGGAGAGACCCCATY 758
265 LeuTYrGlnHisIleLeuTYrPhePheGlyHISProGluValTYrIleI 281
759 CTATTAACCAACCTATCTGATTTTGGTCCACCGAAGTTATATTTCT 808
281 aIleLeuProGlyPheGlyIleIleSerHisValAlaSerThrPheSer 297
809 TATCTTACAGGCTTGGAAATATCTCCATTTGTAACCTTACTCTCGG 858
298 ..LysLYsProValPheGlyTYrLeuProMetValTYrAlaMetValAla 313
859 GAAAAAAGAACCAATTGTGATACATAGTATGCTGACATATGATATCA 908
314 ILeGlyValLeuGlyPheValAlaIleAlaHisMetTYrThrValG 330
909 ATTGATTTCTAGGTTTATCGTGGACACACATATATTTTACAGTAG 958

```

---

```

330 yMetSerLeuThrGlnGlnSerTYrPheMetLeuAlaThrMetValIleA 347
959 AATAGACGTAGACACACAGCATATTTACCTCGGACCATATATCATG 1008
347 laValProThrGlyIleLeuIlePheSerTrpIleAlaThrMetTYrGly 363
1009 CTATCCCAACCGCGGTCAAGATATTATTAGCTGACCTCCACACTCCACGA 1058
364 GlySerValGluPheIleSerProMetLeuTYrPheGlyPheMetPhe 380
1059 AGCATATGAAATGATCTGCTGACGCTCTGACCGCTTGCATTCATC.. 1106
380 eLeuPheThrValGlyValAlaThrGlyIleValIleuAlaGlnIleGlyL 397
1107 .CTTTTACCGGTAGTGGCCCTGACTGCTGCTTGTATTAACCAACTCATCAC 1155
397 euAspArgAlaTYrHISAspThrTYrTYrValAlaIleAlaHisPheHISTYr 413
1156 TAGACATCGTACTACAGACAGCTACTAGCTGTGAGGCCACTTCCACTAT 1205
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTYrPheTY 430
1206 GTCTATCATATGAGACCTGATTTGCCATCATAGGAGGCTTCATTCAC 1255
430 rMetProLYsPheSerGlyArgAlaIlePheProGlyTYrAlaAlaLYsLeuH 447
1256 ATTTCCCTATTTCTCAGGCTACCGCTTACACCAAACTCAGCCAAATCC 1305
447 lAspHeTYrThrPhePheIleGlyAlaAsnValIlePhePheProGlnHIS 463
1306 ATTTACATCATCATATCATCGCGCTAATCTACTTTCTCCCAACAC 1355
464 PheLeuGlyArgGlnGlyMetProArgArgTYrIleAspTYrProGluAl 480
1356 TTTCTGCGCTTATCCGGAATGCCCGACCTTACTCGGACTACCCGATGC 1405
480 aPheAlaLeuTrpAsnLYsValSerSerTYrGlyAlaPheLeuAlaPheA 497
1406 ATACACCAATGAACATCTATCATCTCTAGCTCATTTCTCTGTA 1455
497 laSerPheLeuPhePheIleValIlePheValTYrThrLeuValAlaGly 513
1456 CAGCA...GTAATATTATTAATTTCATGATTTAGAAAGCTTCGCTTCG 1502
514 ArgArgGluThr.....ArgProAsnProTrpGlyIlePheAl 526
1503 AAGCGAAAGTCTTAATAGTAGAAGAACCC.....TC 1534
526 aaSPThrLeuGluTYrThrLeuProSerProProPheAlaHisThrPheG 543
1535 CATTAACCTGAGAGTACTATATGATGATGCCCCCAACCTTACACACATTCG 1584
543 lu 543
1585 AA 1586

```

seq\_name: /SID8/gcgdata/geneseq/geneseg/nm2000.DAT.AA257204

seq\_documentation\_block:

ID: AA257204 standard; DNA; 6691 BP.

AA257204:

30-MAR-2000 (first entry)

Human extramitochondrial DNA sequence SEQ ID NO:3.

Human: mitochondrial DNA; extramitochondrial DNA; mtDNA; exmtDNA;

KW: Huntington's disease; detection; dystonia; Alzheimer's disease;

KW: non-insulin dependent diabetes mellitus; schizophrenia; stroke;

KW: lactic acidosis; myoclonic epilepsy ragged red fibre syndrome;





```

397 euAspATaGAlaTyrHisAspThrTyrTyrValAlaHisPheHisTyr 413
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3522 TAGACATGCTACTACAGACGACTACTACTACTACTACTACTACTACT 3571
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTy 430
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3572 GTCTCATTAATAGACGCTGATTGGCATCATAGAGAGCTCATCAGT 3621
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
430 MetProLysPheSerGlyArgAlaPheProGluTyrPAlaAlaLysLeuH 447
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3622 ATTCCCTCTTCTCAGGCTACACCTTACACCAACCTAGCCCAAAATCC 3671
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
447 IspHerPThrPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3672 ATTTCGCTTTCATATTCATCGGCTAAATCTAATCTTCTCCACACAC 3721
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
464 PheLeuGlyArgGlnGlyMetProArgTyrTyrIleAspTyrProGluAl 480
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3722 TTTCTGGGCTATCCGGAATGCCCGACGTTACTGCGACTATCCGATGC 3771
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
480 aPheAlaLeuThrPasnLysValSerSerTyrGlyAlaPheLeuAlaPhe 497
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3772 ATACACCATGATGAATATCTATCTATGCTAGCGCTCATTTCTCTAA 3821
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
497 IAserPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3822 CAGCAGTAAATATATATTTTTCATATTTGAGAAGCCTTCGCTTCGAG 3871
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
514 ArgAlaGlyThrThrArgProAsnProTyrGlyGluPheAlaAspThrLeu 530
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3872 CGAAAGAGCTTAATAGTAGAA.....GAACCTCCATTAACCTGGA 3912
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
530 uTrrPThrLeuProSerProProAlaHisThrPheGlu 543
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
3913 GTGACTATATGATGCCGCCACCTCTACACATTCCTGAA 3952
    ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: /STD58/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH67734
seq_documentation_block:
ID AAH67734 standard; DNA; 1743 BP.
XX
XX AAH67734;
XX
XX
XX 26-SFP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment seq ID NO: 2769.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX
XX organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX
XX P-PSDB; AAG92515.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX
XX mutation point of a gene, measuring expression of a gene, analysing

```

```

PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 8: SEQ ID NO: 2769; 246bp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 1743 BP; 283 A; 541 C; 427 G; 492 T; 0 other:
S0
XX
XX alignment_scores:
XX      Quality: 1169.50      Length: 556
XX      Ratio: 2.946      Gaps: 11
XX      Percent Similarity: 71.403      Percent Identity: 39.748
XX
XX alignment_block:
XX US-09-712-768-2 x AAH67734
XX
XX Align seg 1/1 to: AAH67734 from: 1 to: 1743
XX
XX 12 HtsGluysGlnGlyPhePheThrArgTyrPheMet.....SerThras 26
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 55 CATCGACGACAGGCG...ACCAAGCATGGTATGATGACACGACCGCA 101
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 26 nHisLysAspIleGlyLeuLeuTyrLeuValAlaIleAlaGlyValAlaGly 43
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 102 CCACACACGCTGGCGCATTTGTACATCATTAATGCTTTCACCTCTTCT 151
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 43 heliSerValLeuPheThrValTyrMetArgLeuGluLeuMetAspPro 59
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 152 TCCCGGTGGCTTGATGGCCCTGCTTATCCGAGCGAGCTTTTCACCCCT 201
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 60 GlyValGlnTyrMetCysLeuGluGlyAlaArgLeuIleAlaAspAlaSe 76
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 202 GGTCTGCGAGTTCTCG.....
XX
XX 76 rGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMetValThrTyrH 93
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 217 .....TCTAATGACGAGTTCACACCGCTGTTCCACCATGC 250
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 93 IsGlyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGly 109
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 251 ACGGACGTGTATGCTGCTGCTGTACGGAACCTCAATTGTGG...GGT 297
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 110 PheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspMetAlaPh 126
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 298 TTTTGCTAATACGCTCGCCACATTCAGATCGTGCGCTGACGTAAGCTTT 347
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 126 eProArgMetAsnAsnLeuSerPheTrpLeuPheIleAlaGlyThrAlaM 143
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 348 CCCACGTTTGAAATGCTTTCGCTTCTGTGATGACACCGCTGCTGTGTCG 397
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 143 etGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySer 159
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 398 CGATGCTGACCGGCTTCTGACCCCGGTGCT.....GCTGCC 435
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 160 GlyValGlyTyrPheValTyrProProLeuSer.....ThrArgGluAl 174
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 436 GACTTGGGTTGACCATGATGCTCCGACATGCTGACGACATTCATCCGCC 485
XX ||||| :||| ||||| ||||| ||||| ||||| |||||
XX 174 aglyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAlas 191
XX ||||| :||| ||||| ||||| ||||| ||||| |||||

```

```

486 AGGCTTGGCTGACATGATGATTTGTCGGTGGTGCATCTGATTTG 535
191 exsrltmetglalaileasmetlethrThpHeuansmetarg 207
536 GCTCCGTTGCTCCGCAATTACATGCTCACCAACATCTCTGCTCCGC 585
208 AlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSerLeu 224
586 GCACCTGGTATGACCATGTTCCGTATGCCATTTTTCACCTGGAAATATCTT 635
224 eilethrAlaTrpLeuLeuLeuAlaLeuProValLeuAlaGlyAla 241
636 CGTGTGTTCCCTTCTCTGCTGATCTTCCACATCTCTGCTGCTGCTG 685
241 lethrMetLeuThrAspArgAsnPhelGlyThrPhePheAsnPro 257
686 CACTGCGTGTCTGTATGACCGCAAGCTGTGGACACCTGTACGATCCA 735
258 AlaGlyGlyGlyAspProIleLeuTrpGlnHisIleLeuTrpPhePhe 274
736 GCTAACGGCGGCTCC...CTCCTGTGGAGACACCTGTTCGTTCTTCG 782
783 ACACCTGATGATTTACGTTCTGCGCTGCTGCTGCTGCTGCTGCTG 832
274 yHisProGlyValTrpIleIleIleLeuProGlyPheGlyIleIleSer 291
291 lValValSerThrPheSerLysLysProValPheGlyTrpLeuProMet 307
833 AGATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 882
308 ValTrpAlaMetValAlaIleGlyValLeuGlyPheValValTrpAla 324
883 ATCTTGCAACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 932
324 shiMetTrpTrpValGlyMetSerLeuThrGlnInsTrpPheMet 341
933 CCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 976
341 euAlaTrpMetValIleAlaValProThrGlyIleLysIlePheSerTrp 357
977 TCATGACGTTCTGATTTGCGTTCTTACCGGCTTAAGTTCTTCAACTGG 1026
358 IleAlaThrMetTrpGlySerValGluPheLysSerProMetLeuTrp 374
1027 GTTGGAACCATGTGGAAGGCTCACATCTTGGGAAACCCCAATGATCTG 1076
374 palArpGlyPheMetPheLeuPheTrpValGlyGlyValThrGlyIle 391
1077 GTCGTGGCTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1126
391 allLeuAlaGlnAlaGlyLeuAspArgAlaTrpHisAspTrpTrpVal 407
1127 TGTGTTGGCTCCGCCACCATGATGCTGCTGCTGCTGCTGCTGCTG 1176
408 ValAlaHisPheHisTrpValMetSerLeuGlyAlaIlePheAlaIle 424
1177 ATCGGCACTTCACTACACCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1226
424 eAlaGlyIleTrpPheTrpMetProLysPheSerGlyArgAlaPhePro 441
1227 TGCAGGCGTTTACTTCTGCTCCGAGATGATGATGCGCGATGATGAG 1276
441 juTrpAlaAlaLysLeuHisPheTrpTrpPheIleGlyAlaAsnVal 457
1277 AGCGTCTTGCAAGATCATTCTGTTGACTTGTGCTGCTGCTGCTGCTG 1326
458 ThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProArgArg 474
1327 ACCTTCTCATCCAGCATGCTGGGGAACATGATGATGCTGCTGCTGCT 1376
474 rIleAspTrp.....ProGluAlaPheAlaLeuTrpAsnLysValSer 489

```

```

1377 CGCTGACTACCTGATTCGATGCTTACCATCTACACAGCATCTCCA 1426
489 exTrpGlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleVal 505
1427 CCGGTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
506 PheValTrpTrpLeuValAlaGlyArgArgGluTrpArgProAsnPro 522
1477 GTCCTCAGTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1526
522 PGLYGLUPHEALASPRHLEUGLUITRPHILEUPROSERPROPROPORA 539
1527 GGGT...TAGCGCAATCCCTGGATGGCAACCTGCGCTCTCTCTC 1573
539 laHisTrpPheGlnTrpLeuProLys.....ArgSerAspTrp 551
1574 GCCACAACTTGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1623
552 AspLysHisProSerHis 557
1624 GAGCTGCACTACCCGAC 1641
seq_name: /STD8/9c9data/geneseq/geneseq/NA2001.DAT:AAH68532
seq_documentation_block:
ID AH68532 standard; DNA: 349980 BP.
AC AAH68532:
DT 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX Corynebacterium glutamicum.
XX EP1108790-A2.
XX 20-JUN-2001.
XX 18-DEC-2000; 2000EP-0127688.
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Corynebacterium, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Disclosure; SEQ ID NO: 7067; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of corynebacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed

```

CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

Sequence 349980 BP; 80900 A; 98397 C; 92139 G; 78544 T; 0 other;

alignment\_scores:

Quality: 1169.50 Length: 556  
 Ratio: 2.946 Gaps: 11  
 Percent Similarity: 71.403 Percent Identity: 39.748

alignment\_block:

US-09-712-768-2 x AAH68532/rev ..

Align seq 1/1 to reverse of: AAH68532 from: 1 to: 349980

```

12 HlaGLuLysGlnGlyPhePheThrArgTrpPheMet.....SerThrAs 26
||| :|||
272751 CATGCACGCAAGGC...AGCAAGCATGTAAATGATGACACACCGCA 272705
26 nHISLysAspIleGlyLeuLeuTyrLeuValAlaIleGlyValAlaGlyP 43
|||||:|||||
272704 CCACAAAGCAGCGGCGATATGATCATCATATATGCTCTTCAAGCTTCT 272655
43 heIleSerValLeuPheThrValTyrMetArgLeuGluLeuMetAspPro 59
||:|||||
272654 TCCTGGGTGCTGATGAGCCCTGCTTATCCGAGCGGAGCTTTTCACCCCT 272605
60 GlyValAlaGlnTyrMetCysLeuGluGlyAlaArgLeuIleAlaAspAla 76
|||||:|||||
272604 GGTCTGCAGTTCCTT..... 272590
76 rGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMetValThrTyrH 93
|||:|||||
272589 .....TCTATGACAGCTTCACACCTGTCACCAATGC 272556
93 iSGLyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGlyIle 109
|||||:|||||
272555 ACGGAACGTGATGCTGCTGCTGACGAACTCCAAATGTTGG...GGT 272509
110 PheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspMetAlaPh 126
|||||:|||||
272508 TTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 272459
126 eProArgMetAsnAsnLeuSerPheThrPheLeuPheIleAlaGlyThrAla 143
|||||:|||||
272458 CCCACGTTTAAATGCTTTCGCTTGTGATCACCACCGTGGGTGGTGG 272409
143 eGlyValAlaSerLeuPheAlaProGlyIleAspGlyLeuGlySer 159
|||||:|||||
272408 CGATGCGACCGGCTTCTGACCCGCGGTGT.....GCTGCC 272371
160 GlyValGlyTyrValLeuTyrProPheLeuSer.....ThrArgGlyAla 174
|||||:|||||
272370 GACTTGGGTGACACATGATACCCACGTCTGACGCGCATTCACCTCCC 272321
174 agIyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAla 191
|||||:|||||
272320 AGGCTTGGCTGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCT 272271
191 eSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsnMetArg 207
|||||:|||||
272270 GCTCCGCTGCTTCCGAAATTACATGCTCACCACCATCTTGCCTGCC 272221
208 AlaProGlyMetThrLeuHisLysValProLeuPheSerTyrSerIlePh 224
|||||:|||||
272220 GCACCTGGATACCATGTCTCCGATGCTTTCACCTGGAATATCTT 272171
224 eIleThrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAlaGlyAla 241
|||||:|||||
272170 CGTGTCTTCCGTTCTTCTGCTGCTGATGATTTCCACCTGCTGCTGCTG 272121
241 leThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePheAsnPro 257

```

```

272120 CACTGGGTCTCTGTATGACCGCAAGCTTGCTGACACCTGTACATCA 272071
|||||:|||||
258 AlaGlyIleGlyAspProIleLeuTyrGlnHisIleLeuThrPhePheG 274
|||||:|||||
272070 GCTAACGCGCGCTCC...CTCCTGTGGCAGACACCTCTCTGCTTCCG 272024
274 yHISProGluValTyrIleIleIleLeuProGlyPheGlyIleIleSer 291
|||||:|||||
272023 ACACCTGAGGTTTACGTTCTGCGCTGCCGCTTCTGCGCATTTCTCT 271974
291 iSValValSerThrPheSerTyrLysProValPheGlyTyrLeuProMet 307
|||||:|||||
271973 AGATGATCTCTGTTCTCCCGTAAGCCCAATGTTGCTGATGCTGGCT 271924
308 ValTyrAlaMetValAlaIleGlyValLeuGlyPheValAlaIlePhe 324
|||||:|||||
271923 ATCTTGCAACCTTGCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTG 271874
324 sHISMetTyrThrValGlyMetSerLeuThrGlnGlnSerTyrPheMet 341
|||||:|||||
271873 CCACATGTTCTGTTACTGGGCGCATTTTCTT.....CCGTTCTTCTCT 271830
341 euAlaThrMetValIleAlaValProThrGlyIleLysIlePheSerTrp 357
||:|||||
271829 TCATGACGCTCTGATTTCCGTTCCACCGCGCTTAAAGTTCTCACTGG 271780
358 IleAlaThrMetTyrPheGlySerValGluPheLysSerProMetLeuTr 374
|||||:|||||
271779 GTTGAACCATGTGAGAGACATACATCTTGGGAAACCCCAATGATCTG 271730
374 PAlaPheGlyPheMetPheLeuPheThrValGlyIleValThrGlyIle 391
|||||:|||||
271729 GTCGTGCTTCATGAGCTTCTCTTCCGCTGCTGCTGCTGCTGCTGCTG 271680
391 aLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyrVal 407
|||||:|||||
271679 TCGTGGCGTCCGCCACACCTGACTTCCACTTGGCTGACTGCTGCTGCT 271630
408 ValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAlaIlePh 424
|||||:|||||
271629 ATGCGACATTCACATACACCTTCTGCTGCTGCTGCTGCTGCTGCTG 271580
424 eAlaGlyIleTyrPheTyrMetProLysPheSerGlyAlaGlyAlaPhePro 441
|||||:|||||
271579 TCGAGCGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271530
441 lUrrAlaAlaLysLeuHisPheTyrThrPhePheIleGlyAlaAsnVal 457
|||||:|||||
271529 AGCGTCTTGCAAGATCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 271480
458 ThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProAlaArgTyr 474
|||||:|||||
271479 ACCTTCTCATCAGCACTGGGTGGGCAACATGGGTATGCAACGCTGTTA 271430
474 rIleAspTyr.....ProGluAlaPheAlaLeuTrpAsnLysValSer 489
|||||:|||||
271429 CGGTGACTGACCTGATGATGATGATGATGATGATGATGATGATGATG 271380
489 eTyrGlyValPhePheLeuAlaPheAlaSerPheLeuPheIleValIle 505
|||||:|||||
271379 CCGTGTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271330
506 PheValTyrThrLeuValAlaGlyArgGlyLutThrArgProAsnProTr 522
|||||:|||||
271329 GTCTTCAAGTCTCTGCGCTGACGAGAGCTGTTACCTGATGATCTCTTG 271280
522 pGlyIlePheAlaAspThrLeuGluTyrThrPheLeuProSerProProAla 539
|||||:|||||
271279 GGGT...TACGGCACTCTCCGAGTGGGCAACCTCTGCGCTCTCTCTC 271233
539 lAHisThrPheGluThrLeuProLys.....ArgSerAspTrp 551
|||||:|||||

```

271232 GCCACAACTTGCATCTGCTCGATCCGCTCCGAGCCCTGCTTC 271183

552 AsplysHisProSerHis 557

271182 GAGCTGCCTACTACCCGAC 271165

seq\_name: /SID58/gcgdelta/geneseq/geneseqn/NA2001.DAT:AAF71686

seq\_documentation\_block:

ID AAF71686 standard; DNA: 1779 BP.

AA71686;

30-APR-2001 (first entry)

Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:653.

Corynebacterium glutamicum; carbon metabolism and energy production;  
SMP protein; sugar metabolism and oxidative phosphorylation protein;  
fine chemical production; organic acid; proteinogenic amino acid;  
nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
diagnosis; Corynebacterium diptheriae; evolutionary study; ds.

Corynebacterium glutamicum.

WO200100844-A2.

04-JAN-2001.

23-JUN-2000: 2000MO-1B00943.

25-JUN-1999: 99US-0141031.

08-JUL-1999: 99DE-1031412.

08-JUL-1999: 99DE-1031413.

08-JUL-1999: 99DE-1031419.

08-JUL-1999: 99DE-1031420.

08-JUL-1999: 99DE-1031424.

08-JUL-1999: 99DE-1031428.

08-JUL-1999: 99DE-1031431.

08-JUL-1999: 99DE-1031433.

08-JUL-1999: 99DE-1031434.

08-JUL-1999: 99DE-1031510.

08-JUL-1999: 99DE-1031562.

08-JUL-1999: 99DE-1031634.

09-JUL-1999: 99DE-1032180.

09-JUL-1999: 99DE-1032227.

09-JUL-1999: 99DE-1032230.

09-JUL-1999: 99US-0143208.

14-JUL-1999: 99DE-1032924.

14-JUL-1999: 99DE-1032973.

14-JUL-1999: 99DE-1033005.

27-AUG-1999: 99DE-1040765.

31-AUG-1999: 99US-0151572.

03-SEP-1999: 99DE-1042076.

03-SEP-1999: 99DE-1042079.

03-SEP-1999: 99DE-1042086.

03-SEP-1999: 99DE-1042087.

03-SEP-1999: 99DE-1042088.

03-SEP-1999: 99DE-1042095.

03-SEP-1999: 99DE-1042123.

03-SEP-1999: 99DE-1042125.

(BADI) BASF AG.

Pompejus M, Kroegeer B, Schroeder H, Zeider O, Haberhauer G;

WPI: 2001-061975/07.

P-PSDB: AAF79569.

New isolated Corynebacterium glutamicum nucleic acid encoding a sugar metabolism and oxidative phosphorylation protein for production or

modulation of production of fine chemicals e.g. amino acids, carbohydrates or enzymes -

Claim 3: Page 1069-1072; 1246pp; English.

AA71360 to AAF71750 encode the Corynebacterium glutamicum sugar metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243 to AAB 79633 which are involved in carbon metabolism and energy production. The C. glutamicum SMP gene can be used in vectors (II) for expression in host cells and production or modulation of production of fine chemicals, such as, an organic acid, a proteinogenic or nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins (II) encoded by them are used for diagnosing the presence or activity of Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells containing them are used to map genomes of organisms related to C. glutamicum, identify and localise C. glutamicum sequences of interest, in evolutionary studies, in determining SMP protein regions required for function, in modulating SMP protein activity, in modulating the metabolism of sugars, and in modulating high-energy molecule production in a cell (i.e. ATP, NADPH).

Sequence 1779 BP; 294 A; 548 C; 435 G; 502 T; 0 other:

alignment\_scores: Quality: 1166.50 Length: 556

Ratio: 2.946 Gaps: 11

Percent Similarity: 71.223 Percent Identity: 39.748

alignment\_block:

US-09-712-768-2 x AAF71686

Align seg 1/1 to: AAF71686 from: 1 to: 1779

12 HisGluysGinglyPhepethrAtrpPheMet.....SerThras 26

68 CATGCACGCCAGGGC...AGCAAGCATGGTTATGTATGATGCACCAACCA 114

26 nHisLysAsp1LeuLeuLeuTyrlLeuValAlaAlaGlyValGlyP 43

115 CCACAGCAGCGCTGGCATATATATATATATATATATATATATATATAT 164

43 heLleSerValLeuPhePheValTyrMetArGleuGluLeuMetAspPro 59

165 TCCTGGTGCGCTTGTATGGCCTGCTTATCGAGGAGGAGCTTTACCCCT 214

60 GlyValGlnTyrMetCysLeuGluGlyAlaArgLeuIleAlaAspAla 76

215 GCGTGCAGATTCCTG..... 229

76 rGlnThrCysThrAlaAsnGlyHisLeuTyrPAsnValMetValThrTyr 93

230 .....TCTAATGACAGCTTCACACAGCGTTCACCAATGC 263

93 IsGlyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGly 109

264 ACGGACGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 310

110 PheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspMetAlaPh 126

311 TTTCCTAAGTACGTCCTGCGACCTGATGATGATGATGATGATGATGAT 360

126 eProArgMetAsnAsnLeuSerPheThrPhePheIleAlaGlyThrAla 143

361 CCGACGTGATGATGCTTCCGCTTCCGATACACACCGCTGCTGCTGCTG 410

143 etGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnIleGlySer 159

411 CGATGCTGACCGGCTTCCGACCCGCGGTGT.....GCTGCC 448



CC from Coryneform bacterium, and identifying a homologue of a gene deriving  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

aa sequence 1596 BP; 469 A; 384 C; 492 G; 251 T; 0 other

alignment_scores:	Length:	542
Quality:	Gaps:	5
Ratio:		2.982
Percent Similarity:	Percent Identity:	40.03
		71.956

alignment\_block:  
US-09-712-768-2 x AAH67733/rev

Align seg 1/1 to reverse of: AAH67733 from: 1 to: 1596

[illegible][illegible]



536 OProProAlaHisThrPheGluThrLeuProlys.....Args 549  
 115 TTTCTCTCGCCACACTTCGCACTCTTCCCTGTATCCGCTCCGACGCC 66  
 549 eAspTrpAspLysHisProSerHis 557  
 65 CTGCGTTTCGAGTCGACTACCCGAC 40

seq\_name: /SID8/gcgdata/geneseq/geneseq/NA2001.DAT:AAH52833

seq\_documentation\_block:

ID AAH52833 standard; DNA; 1947 BP.

AC AAH52833;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1059.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;

OS Staphylococcus epidermidis.

PN MO20134809-A2.

PD 17-MAY-2001.

PE 09-NOV-2000; 2000MO-US30782.

PR 09-NOV-1999; 9905-0164258.

PA (GLAXO) GLAXO GROUP LTD.

PI Kimerly Wf;

DR WPI: 2001-316495/33.

DR P-PADB: AAG81983.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

PT useful for vaccinating against infections, e.g. endocarditis.

PS Claim 8: Page 308-309; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention. N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 1947 BP; 538 A; 367 C; 361 G; 681 T; 0 other;

alignment\_scores:

quality: 1078.50 length: 550  
 Ratio: 2.794 Gaps: 8  
 Percent Similarity: 70.182 Percent Identity: 38.545

alignment\_block:

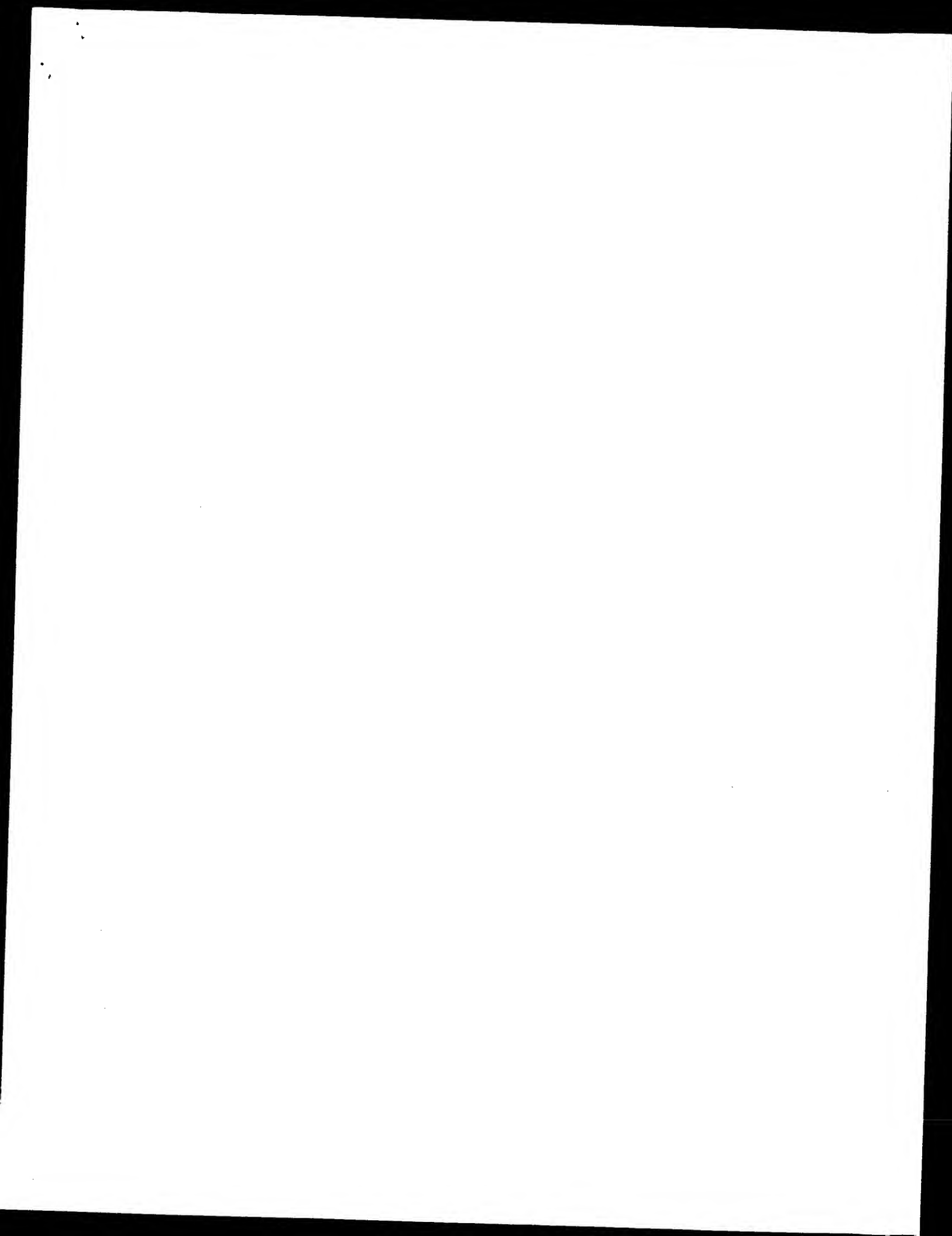
US-09-712-768-2 x AAH52833 ..

Align seg 1/1 to: AAH52833 from: 1 to: 1947

18 PheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeuLeu 34  
 88 TATAAGAGATGTTTACATCCGTAGACCAATAAATAATCGTATCATGTA 137  
 34 rLeuValAlaAlaGlyValAlaGlyPheIleSerValLeuPheThrVal 51  
 138 TTTAATTTCTGCCGTATTAAATGTTCTGTCGTGTCATCATGCTTAA 187  
 51 yrMetArgLeuGluLeuMetAspProGlyValGlnTyrMetCysLeu 67  
 188 TGTTACGTACTCAATTAACAAATTCAGATACAAATCTTCG..... 228  
 68 GlyAlaArgLeuIleAlaAspAlaSerGlnThrAlaAsnGlyH 84  
 229 .....GAGGCAA 236  
 84 sLeuTrpAsnValMetValThrTyrHisGlyIleLeuMetPhePhe 101  
 237 CCACCTATTAATGAGATATTACTACGCGCGTAATATGATATTATTA 286  
 101 aGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 117  
 287 TGCTATGCGCATTTATCTTT...GTTATGCGATGTTGTTATTCATTA 333  
 118 GlnIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeuSer 134  
 334 CAACCTGGTGCACGCGATGTTCCCTCCGTAAATGATATACCTAGTAT 383  
 134 eTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAla 151  
 384 CTGGCTATTCTCGCTGATGATTTATTCACACCTATCA...TTTATG 430  
 151 roGlyGlyAspGlyGlnLeuGlySerGlyValGlyTyrValLeuTyr 167  
 431 TAGTGGA.....TCACACGCTGCTGGTGGACCTACACGCA 468  
 168 ProLeuSerThrArg...GlnAlaGlyTyrSerMetAspLeuAlaIle 183  
 469 CCACCTGCTGCTGAATTCAGTCCAGGTCACGAGTCACTATATTATTA 518  
 183 eAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMet 200  
 519 TGCAATTCAAATATCTGGTATCGGATCGTTAAATGATCGTATCACT 568  
 200 l eThrTrpPheLeuAsnMetArgAlaProGlyMetThrLeuHisLysVal 216  
 569 TTGTACGATTTAAGATGTAACACCCACAAATGAATTTATGCAATG 618  
 217 ProLeuSerTrpSerIlePheIleThrAlaTrpLeuIleLeuAla 233  
 619 CCAATGTTCAAGTAAACAACATTCATTAACATTAATGTTATATGAC 668  
 233 aLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArg 250  
 669 ATCCAGTGTTCACGTAGACCTGCTTATATGACGTGATGAAATTT 718  
 250 heGlyThrThrPhePheAsnProAlaGlyGlyAspProIleLeuTyr 266  
 719 TTGGACGTCAGTTCTTACGTAGCAATGCGGTATGCCAATGCTTTG 768  
 267 GlnHisIleLeuTrpPheGlyHisProGlyValTyrIleIleIle 283  
 769 GCAACTCTTCTGGGTATGGGCGCACCCCTGAAGTTATATCGTATTT 818  
 283 uProGlyPheGlyIleIleSerHisValValSerThrPheSerLysLeu 300  
 819 GCCAGCATTCGTATGACTCAGAAATCATCCCTACTTTGCCCGTAAC 868



300 rovalPheGlyTyrLeuProMetValTyrAlaMetValAlaIleGlyVal 316  
:::|||||::: |||::: |||::: |||:::  
869 GTTATTGCGTCATCAAGTATGATTTGGCACTGCAGATGCGATTC 918  
317 LeuGlyPheValValTyrPAlaHisIleMetTyrThrValGlyMetSerLe 333  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
919 TTAACTTCTTAGTTGGTTCACCAATTCCTACATAGGTAATGGTGC 968  
333 uThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValProT 350  
::: |||::: |||::: |||::: |||::: |||::: |||:::  
969 GTTAATTAACTCATTTCTCTCATCTCAACAATGTTAATCGGTCCAA 1018  
350 hrGlyIleuIlePheSerTyrPAlaIleThrMetPrglyGlySerVal 366  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1019 CCGAGTTAACTAATTTAACTGCTGCTCCACATTTATACAAAGTAGAAT 1068  
367 GluPheIysSerProMetLeuThrPAlaPheGlyPheMetPheLeuPheTh 383  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1069 ACATTTGAGTCACCTATGCTATTCATAGCATTCATCCCTAATCTTCT 1118  
383 rValGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAspArg 400  
::: |||::: |||::: |||::: |||::: |||::: |||:::  
1119 AFTAGAGAGGCTACTGCTGTAATGCTGCAATGCGATCAGCTGACTATC 1168  
400 IaTyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMetSer 416  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1169 AATATCACAACACTTATTTCTTACTACTCCTCCACTATACATTCGTT 1218  
417 LeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetProly 433  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1219 ACTGGTAGATTTGCTGCTGCTAGCTGTTAATCTTCTGCTATCCAA 1268  
433 sPheSerGlyArgAlaPheProGluTyrPAlaAlaIleuHisPheTyrP 450  
::: |||::: |||::: |||::: |||::: |||::: |||:::  
1269 AATGATGGCTACAACTTAATGAACACTTAACAACATGGCTTCGTG 1318  
450 hrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuGly 466  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1319 TCTTCATGATCGATTAACTTCTTCTTACCAATTCATTCATTCAGT 1368  
467 ArgGlnGlyMetProArgArgTyrIleAspTyrProGluAlaPheAlaLe 483  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1369 TTGATGATGATGCCAGCTGCTCTATACACTTACCTCTGATGCTTG 1418  
483 uTP.....AsnIysValSerSerTyrGlyAlaPheLeuAlaPheAlas 498  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1419 GTGGTTACTAAGCTCTCACTGATCGTGCAGTATTCATGATGCAATTG 1468  
498 erPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGlyArg 514  
::: |||::: |||::: |||::: |||::: |||::: |||:::  
1469 GATTCCTAATTCCTAGTTGCAAGTATGCTTATTCATATCAAAAGCTCCA 1518  
515 ArgGluThrArgProAsnProTyrGlyGluPheAlaAspThrLeuGluTr 531  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1519 CGTGAAGCTACTGAGATTAAGTGCAGTGGTGCCTTACTTTAGAAAT 1568  
531 rThrLeuProSerPro...ProProAlaHisThrPheGluThrLeuProL 547  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1569 GCTACAGCATCAGCTATTCACCTAAATACAACTTTCATACACCTCCTG 1618  
547 ysArgSerAspTyrAsp.....LysHisProSerHis 557  
|||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1619 ATTGGAAATGATACATTCATTCGTTGATATGAAGAACAATGGTCTCAT 1668











```

seq.name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-097-889-2
; Sequence 2, Application US/09097889
; Patent No. 6218117
GENERAL INFORMATION:
APPLICANT: Herrnstadt, Corrina
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Davis, Robert E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,889
FILING DATE: 15-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Roseman Ph. D., Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 660088.417
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-6031
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16569 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
US-09-097-889-2

alignment_scores:
Quality: 1328.00 Length: 534
Ratio: 3.192 Gaps: 7
Percent Similarity: 77.903 Percent Identity: 50.187

alignment_block:
US-09-712-768-2 x US-09-097-889-2 ..

Align seg 1/1 to: US-09-097-889-2 from: 1 to: 16569

17 phephethargtrphenectserthrasnhslsyaapileglyleu 33
||||| : : : : : |||
5907 TTCGCGCGACCTGCATATCTCTACAAACCAAGACATTTGACACT 5956
33 uTyrLeuValAlaIaGlyValAlaGlyPheIleSerValLeuPheThrv 50
||||| : : : : : |||
5957 ATACTATTATTTCGGCGCATGACCTGGATCCTAGACACACTTAAGCC 6006
50 aTyrMetArgLeuGluLeuMetAspProGlyValGlnTyrMetCysLeu 66
:: : : : ||| : : : |||
6007 TCCTATTTCGAGCGACCTGCGGACCGACGAGC..... 6038
67 GluGlyAlaArgLeuIleAlaIaSerPalaSerGlnThrCysThrAlaAsnG 83
:: : : : |||
6039 ..... AACCTTCTAGTAACGA 6055

```

```

83 yHisLeuTrpAsnValMetValThrHisGlyIleLeuMetMetPheP 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6056 CCACATCTACACGCTATCTCTACACGCCATGCGATTGTAATATCTTCT 6105
100 heValGlyIleProAlaLeuPheGlyIlePheGlyAsnTyrLeuMetPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6106 TCATAGTAATACCATCATATGAGGCTTTGGCAACTGACATGCTTCCC 6155
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeu 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6156 CTATATATAGTGTGCCCCGATATGGGCTTTCCCGCATTAACAACTAG 6205
133 rPheTyrPhePheIleAlaGlyThrAlaMetGlyAlaSerLeuPheA 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6206 CTTCCTACTCTTACTCTCTCTCTCTACTCTCTGCTGCTGCTATAG 6255
150 IapProGlyIleAspGlyIleLeuGlySerGlyValGlyTyrValLeuTyr 166
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6256 TG.....GAGGCCGAGACGGAACAGGTTGAACAGCTTAC 6290
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6291 CTTCCCTTAGACAGGAACTACTCCACCCCTGAGGCTCCGTAAGACTTAC 6340
181 alIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleA 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6341 CATCTTCTCTTACACCTAGACAGGCTCTCTCTATCTTACGGGCTATCA 6390
198 smetIleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHis 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6391 ATTTCATGACACACATATATCAATATAAACCCCTGCCATTAACCAATAC 6440
215 LysValProLeuPheSerThrSerIlePheIleThrAlaTyrPheLeuIle 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6441 CAACGCCCTCTCTCTGCTGATCCGCTTATATACACAGCTTACTTCT 6490
231 uLeuAlaLeuProValIleAlaGlyAlaIleThrMetLeuLeuThrAspA 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6491 COTATCTCTCCAGTCTTACGCTGCTGCTGCTATCTACTACTAACACACC 6540
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyIleGlyAspProIle 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6541 GCAACCTTCACACACCTTCTTCGACCCCGGAGAGAGACCCCAT 6590
265 LeuTyrGlnHisIleLeuTrpPhePheGlyHisProGlyValTyrIleI 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6591 CTATACCAACACCTATCTGATTTTCTGTCACCCGTAAGCTTATATCT 6640
281 eIleLeuProGlyPheGlyIleIleSerHisValIleSerThrPheSer. 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6641 TATCTCTACGAGGCTTCGGAATATCTCCCATATTTGTAACCTACTCTCCG 6690
298 .LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6691 GAAAAAAGAACCATTTGGATACATAGTATGCTGCTGACCTATGATATCA 6740
314 IlleGlyValLeuGlyPheValIleTrrAlaHisIleMetTyrThrValG 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6741 ATTGGCTTCTAGGTTTATCGTGTAGACACACATATTTTACAGTATAGG 6790
330 yMetSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleA 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6791 AATAGAGGTAGACACAGCATATTTTCACCTCCGCTACCATATATCAG 6840
347 IValAlaProThrGlyIleLysIlePheSerTrrIleAlaThrMetTrrPgly 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6841 CTATCCCAACGCGGCTCAAGTAATTTAGCTGACTGCCACACACTCACGGA 6890
364 GlySerValGluPheLysSerProMetLeuTrrAlaPheGlyPheMetPh 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6891 AGCAATATGAAATGATCTGCTGACAGTCTGAGCCCTAGAGTTTATCTT 6940

```

```

380 eLeuPheThrValGlyValThrGlyIleValIleAlaGlnAlaGlyI 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6941 TCTTTTACCGGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6990
397 euAspArgAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyr 413
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6991 TTAACATCTGATCTACACGACACCTACTACTGTTGAGCCCATCTTCCAT 7040
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyr 430
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7041 GTCTTTCATATAGACTGATATTTGCTATCATATAGGAGCTTCAATTCAG 7090
430 rMetProLysPheSerGlyArgAlaPheProGlyTrrAlaAlaLysLeuH 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7091 ATTTCCTATTTCTGAGCTTACACCTGACCAACCAACCTAGCCAAATCC 7140
447 iAspThrPheThrPheIleGlyAlaAsnValThrPhePheProGlnHis 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7141 ATTTCATCTATCATTTATTCATGCGGCTAAATCTTACTTCTTCCACAC 7190
464 PheLeuGlyArgGlnGlyMetProArgTrrGlyIleAspTrrProGlyAl 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7191 TTTCTGCGGCTTATCCGGAATGCGCGAGCTTACTGAGACTACCCGATCC 7240
480 aPheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheA 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7241 ATACACACATGAAACATCTATCATCTGTAGGCTCATCTTCTCTGTA 7290
497 lAsSerPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7291 CAGCA...GTATATATATATTTTTCATGATTTAGAGAGCTTCCGCTTG 7337
514 ArgArgGluThr.....ArgProAsnProTrrPglyLupheAl 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7338 AAGCAAAAGTCTAATAGTAGAAGAACCC.....TC 7369
526 aAspThrLeuGlnIleTrrPheLeuProSerProProAlaHisThrPheG 543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7370 CATTAACCTGAGTACTATATGATGATGCTCCCAACCTTACCACACATTCG 7419
543 Lu 543
||
7420 AA 7421
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-413-740A-1
seq_documentation_block:
; Sequence 1, Application US/08413740A
; Patent No. 6171859
; GENERAL INFORMATION:
; APPLICANT: HERRNSTADT, CORINNA
; APPLICANT: PARKER, WILLIAM D.
; APPLICANT: DAVIS, ROBERT
; APPLICANT: MILLER, SCOTT W.
; TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
; TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondria
; NUMBER OF SEQUENCES: 206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036-5405
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,740A
; FILING DATE:

```





```

1356 TTCTCGGCGCATCCGGAATGCCCGGAGCTTACGAGCTACACCCGATGC 1405
480 apheAlaLeuTrpAsnValSerSerTyrGlyAlaPheLeuAlaPhe 497
1406 ATACACCAATGAACATCATCTATCTGTAGGCTCATCTTCTCTAA 1455
497 LaSerPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
1456 CAGCA...GTAATATATATATATTTCAATGATTTGAGAAGCCTTCGCTTCG 1502
514 ArgArgGlyThr.....ArgProAsnProTrpGlyGluPheAl 526
1503 AAGGGAAGAGCTTAATAGTAGAAGAACCC.....TC 1534
526 aAspThrLeuGluTrpThrLeuProSerProProAlaHisThrPheG 543
1535 CATAAAGCTGAGTACTATATGATGATGCCCGCCACCTACACACATTCG 1584
543 Lu 543
1585 AA 1586

```

seq\_name: /cgn2\_6/ptodata/2/ina/PCUS\_COMB.seq:PCT-US95-04063-1

```

seq_documentation_block:
; Sequence 1, Application PC/TUS9504063
; GENERAL INFORMATION:
; APPLICANT: HERRNSTADT, CORINNA
; APPLICANT: PARKER, WILLIAM D.
; APPLICANT: DAVIS, ROBERT W.
; APPLICANT: MILLER, SCOTT W.
; TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
; TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
; NUMBER OF INVENTION: Defects
; NUMBER OF SEQUENCES: 206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036-5405
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04063
; FILING DATE: 30-Mar-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonham, David B.
; REGISTRATION NUMBER: 34297
; REFERENCE/DOCKET NUMBER: 2105/7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-1776
; TELEFAX: (202) 429-0796
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-04063-1

```

alignment\_scores: Quality: 1311.50 Length: 534  
Ratio: 3.160 Gaps: 8

Percent Similarity: 77.715 Percent Identity: 50.000

alignment\_block:  
US-09-712-768-2 x PCT-US95-04063-1 ..

Align seg 1/1 to: PCT-US95-04063-1 from: 1 to: 1735

```

17 PhePheThrArgTrpPheMetSerThrAsnHisValAspIleGlyLeuLeu 33
75 TTGCGCGACCGTTGACTATTCTCTACAAACACAAAGACATTTGAGAACT 124
33 uTyrLeuValAlaAlaGlyValValGlyPheIleSerValLeuPheTrp 50
125 ATACCTATTATTTCGGCGCATGAGCTGAGCTCCTAGGACAGCTTAAGCC 174
50 aTyrMetArgLeuGluLeuMetAspProGlyValGlyIleTyrMetCysLeu 66
175 TCCTTATTCGAGCGGAGCTGGGCGAGCCAGGC..... 206
67 GluGlyAlaArgLeuLeuAlaAspAlaSerGlnTrpCysThrAlaAsnG 83
207 .....AACCTTCTAGGTAAACA 223
83 yHisLeuTrpAsnValMetValThrTyrHisGlyIleLeuMetMetPhe 100
224 CCACATCTACAAAGCTATATGCTACAGCCCATGCTATTGTAATATCTTCT 273
100 heValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 116
274 TCATAGTAATACCCATCAATATGAGGCTTTGGCAACTGATTAAGTCC 323
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeuSe 133
324 CTAATATATCGTGCCCGCATATGCGGCTTCCCGCATTAACAAACATATAG 373
133 rPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheA 150
374 CTCTGACTCTTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAG 423
150 LaProGlyLysAspGlyGlnLeuGlySerGlyValGlyTyrValLeuTyr 166
424 TG.....GAGCGCGAGCGAGNACAGGTTGAACAGCTTAC 458
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181
459 CCTCCCTTAGCAGGAGACTATCCACCCCTGGAGCTCCGTAGACTTAC 508
181 alIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleA 198
509 CATCTTCTCTTACACCTAGCAGGCTCTCTCTCTCTCTCTCTCTCTCTCT 558
198 smMetIleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuLeu 214
559 ATTTCATCCACACATTAATATATAAAACCCCTGCGCATTAACCAATATAC 608
215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLe 231
609 CAACGCGCCCTCTCTGCTGATCGCTGCTCAATACACAGCAGCTTACTCT 658
231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuTrpAsp 248
659 CCTATCTCTCCAGCTCAGCTGAGCTGAGCTACATTAAGTAAAGAGAC 708
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAlaAspProIle 264
709 GCACCTCAACACACCTTCTTTCGAGCCCGGAGGAGAGACCCATTT 758
265 LeuTyrGlnHisIleLeuTrpPhePheGlyHisProGlyValTyrIleI 281
759 CTAATCAACACCTATCTGATTTTTCGTCACCCCTGAAGTTATATCTCT 808
281 eIleLeuProGlyPheGlyIleIleSerHisValValSerThrPheSer 297

```

```

809 TATCTACACAGGCTTGGAATATCTCCATATTTGTAATCTACTACTCCG 858
298 ..LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
859 GAAAAAAGAACCATTTGGATACATAGTATGCTGACACTATGATATCA 908
314 TLeGlyValLeuGlyPheValValTyrAlaHisHisMetTyrThrValG1 330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
909 ATTGGATTCCTAGGCTTTATCTGCTGACACACACATATATTACAGTAGG 958
330 yMetSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleA 347
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
959 AATAGACGTAGACACAGACATATTTACCTCCGCTACCATATATCATG 1008
347 IValAlaProThrGlyIleLysIlePheSerTyrPheAlaThrMetTyrGly 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1009 CTATCCCAACCGCGGTCAAGATTTAGTACTGCCCACTCCAGCA 1058
364 GlySerValGluPheLysSerProMetLeuTyrPalaPheGlyPheMetPn 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1059 AGCAATATGAAATGATCTGCTGAGTGTCTGAGCCCTAGATTCTATC.. 1106
380 euePheThrValGlyValTyrGlyIleValIleAlaGlnAlaGlyL 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1107 .CTTTTCACCGTAGGTGGCTGACGTGGCATTTGATTAAGCAATCATC 1155
397 euAsparGalaTyrHisAspThrTyrTyrValValAlaHisPheHisTyr 413
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1156 TAGACATCTGACTACAGACAGACGACTAGTGTGAGCCACATTCACAT 1205
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTy 430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1206 GTTCCTATCATAGAGCTGATTTGGCATCAATAGAGGCTTCATTCAC 1255
430 rMetProLysPheSerGlyArgAlaPheProGluTyrPalaIleLysLeu 447
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1256 ATTTCCTATTTCTCAGGCTACACCTAGACCAACCTACGCCAAATCC 1305
447 IsPheThrPhePhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1306 ATTTCACATCATATTCATCGGGTAAATCTAATCTTCTCCACACACAC 1355
464 PheLeuGlyArgGlnGlyMetProArgTyrIleAspTyrProGlnAl 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1356 TTTCTCGGCTATCCGGAATGCCCGACCTTACCTGACCTACCCGATGC 1405
480 apPheAlaLeuTyrPasnLysValSerSerTyrGlyAlaPheLeuAlaPheA 497
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1406 ATACACCATGAAACATCTCATCTGTAGGCTCATTTCTCTCTA 1455
497 IsSerPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1456 CAGCA...GTAATATTAATTAATTTTCATGATTTAGAAAGCTTCGCTCG 1502
514 ArgArgGluThr.....ArgProAsnProTyrPglyGluPheAl 526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1503 AAGCGAAAGCTCTAATAGTAGAAGACCC.....TC 1534
526 aAspThrLeuGluTyrThrLeuProSerProProAlaHisThrPheG 543
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1535 CATAAACCTGGAGTGTATGATGATGCCCCACCTCTACACACATTCG 1584
543 Lu 543
      ||
1585 AA 1586

```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.us-09-097-889-1

seq\_documentation\_block:

; Sequence 1, Application US/09097889

; Patent No. 6218117

; GENERAL INFORMATION:

; APPLICANT: Herrnstadt, Corrina

```

; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,889
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.417
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-097-889-1

```

```

alignment_scores:
  Quality: 1289.00      Length: 531
  Ratio: 3.144          Gaps: 6
  Percent Similarity: 77.213  Percent Identity: 49.718

```

alignment\_block:  
US-09-712-768-2 x US-09-097-889-1 ..

Align seg 1/1 to: US-09-097-889-1 from: 1 to: 6744

```

17 PhePheThrArgTyrPheMetSerThrAsnHisLysAspIleGlyLeu 33
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2481 TTCCGCCGACCGTGTGATCTCTTACAAACCAACGACATTTGGAACAT 2530
33 uTyrLeuValAlaIleGlyValIleGlyPheIleSerValLeuPheThrV 50
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2531 ATACCTATTATTCGGCGATGAGCTGGAGTCTTAGGCACACGCTTAAGCC 2580
50 aTyrMetArgLeuGluLeuMetAspProGlyValGlnTyrMetCysLeu 66
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2581 TCCTATTTCGAGCGGACGARTGGGCGACGACGCG..... 2612
67 GluGlyAlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnG1 83
2613 .....AACCTCTAGTAGCA 2629
83 yHisLeuThrPasnValMetValThrTyrHisGlyIleLeuMetMetPhe 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2630 CCACATCTTACACAGTATTCGTCACACCCCGCATGCTTGTGATATATCTTCT 2679
100 heValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 116
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2680 TCATAGTATATCCCATCATATATCGAGGCTTTGGCACTGACTGATGTTCC 2729
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeuSe 133

```



```

alignment_scores:      Quality: 487.50      Length: 231
                        Ratio: 3.066      Gaps: 1
Percent Similarity:    68.831      Percent Identity: 41.5556
alignment_block:
  US-09-712-768-2 x US-08-998-416-1138  ..

```

Align seg 1/1 to: US-08-998-416-1138 from: 1 to: 719

```

333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPr 349
4 CTATATCTAGTACAGCTTATTTTACTGCTACATATATATCTGTTCTCC 53
349 oThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGlySerV 366
54 TACTAGATTAATAAGTATTTGATGATACATATTTATGTTGGTTCAT 103
366 aGluPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPhe 382
104 TAGATTAAGTACACCAATATATATCTATATCATTTTATTTTATTT 153
383 ThrValGlyIleValThrGlyIleValLeuAlaGlnAlaGlyLeuAsp 399
154 ACGTAGTGGTGTAACTGCTAGTATGATTAAGCTATCTATGATGATCT 203
399 gAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMet 416
204 AGCATTCATGATACATTTATATGATGATACATTTCCATTTATGATTA 253
416 eLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetPro 432
254 GTTAGTGTCTGATCTGATGCTGATGCTGATGCTGATGATGATGATGCT 303
433 LysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTr 449
304 CTGTGTTTGGTTTAAATGATATGAAATGATGATGATGATGATGATGCT 353
449 pThrPhePheIleGlyAlaAspValThrPhePheProGlnHisPheLeu 466
354 ATTAATTTTCTTATGCTTATATATATGTTTCTCCATGATGATGATGCT 403
466 LysGlnGlnIleMetProArgArgTyrIleAspTyrProGlnAlaPheAla 482
404 GTATTATATGCTTATGATGATGATGATGATGATGATGATGATGATGCT 453
483 LeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPhe 499
454 GGTTCATATTTAGTATCTGATGATGATGATGATGATGATGATGATGAT 503
499 eLeuPhePheIleValIlePheValTyrThrLeuValAlaGlyArgArg 516
504 AATGTTATTTCTTATATTTATTTATGATGATGATGATGATGATGATGAT 553
516 LysThrArgProAsnProTrpGlyLeuPhe..... 525
554 ATAAAGTTAATTAATTAATCTATTAATTAATTAATTAATTAATTAAT 603
526 ..... AlaAspThrLeuGln 530
604 GAATCAATATATTTCTTATGATGATGATGATGATGATGATGATGATGAT 653
530 uTrpThrLeuProSerProProAlaHisThrPheGluThr 544
654 GTTATATTAATTAATTAATCACTTATTAATTAATTAATTAATTAAT 696

```

seq\_name: /cogn2\_6/prodata/2/lna/6B\_COMB.seq:US-08-998-416-187

seq\_documentation\_block:

Sequence 187, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jürgen

APPLICANT: Knechtle, Philipp

APPLICANT: Reibischung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYPTI

```

TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG10740P
US-08-998-416-187

```

alignment\_scores:

Quality:	Ratio:	Length:
470.00	3.357	181
Percent Similarity: 77.348	Percent Identity: 48.066	Gaps: 0

alignment\_block:

US-09-712-768-2 x US-08-998-416-187 ..

```

Align seg 1/1 to: US-08-998-416-187 from: 1 to: 663
333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPr 349
4 CTATATCTAGTACAGCTTATTTTACTGCTACATATATATTTTCTTATTC 53
349 oThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGlySerV 366
54 TACTAGATTAATAAGTATTTGATGATACATATTTATGTTGGTTCAT 103
366 aGluPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPhe 382
104 TAGATTAAGTACACCAATATATATCTATATCATTTTATTTTATTT 153
383 ThrValGlyIleValThrGlyIleValLeuAlaGlnAlaGlyLeuAsp 399
154 ACTAGTGTGGTTTAACTGCTAGTATGATTAAGTAACTGATTCATTA 203
399 gAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMet 416
204 AGCATTCATGATACATTTATATGATGATACATTTCCATTTATGATTA 253
416 eLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetPro 432
254 GTTAGTGTCTGATCTGATGCTGATGCTGATGCTGATGATGATGATGCT 303

```

```

433 LysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTr 449
      |||      ::| |||      ::::::::::::::|
304 CTGTTTAAAGTTAAATTAATGAAGAAATTAATCACAATTCATTCG 353
      ::::::::::| |::| |::| |::| |::| |::| |
449 PThrPheHeliGlyAlaAsnValThrPhePheProGluHisPheLeuG 466
      ::::::::::| |::| |::| |::| |::| |::| |
354 ATTAATTTCTAGTCTTAATTAATTTCTCTCCATGATTCCTAG 403
      ::::::::::| |::| |::| |::| |::| |::| |
466 LysArgGlnGlyMetProArgArgTyrIleAspTyrProGluAlaPheAla 482
      ||| ::::| |::| |::| |::| |::| |::| |
404 GTATTAATGATACCAAGAAATTCCTGATTAATCCGATCATTCCTA 453
      ||| ||||| |::| |::| |::| |::| |::| |
483 LeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPh 499
      ||| ||||| |::| |::| |::| |::| |::| |
454 GGTGAATTTAGTATCTTCATTTGGTTCCTATTAATCACTATATCATTT 503
      ::::::::::| |::| |::| |::| |::| |::| |
499 eLeuPheHeliValIlePheValTyrThrLeuValAlaGly 513
      ::::::::::| |::| |::| |::| |::| |::| |
504 AATGTTATCTTATATTATTATGATCAATTAATAATGTT 546
      ::::::::::| |::| |::| |::| |::| |::| |

```

seq\_name: /cgn2\_6/protdata/2/lna/6B\_COMB.seq:US-08-998-416-289

seq\_documentation\_block:

```

; Sequence 289, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtel, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwalis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 289:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 856 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORGANISM: PAG1241UP
; US-08-998-416-289

```

# alignment\_scores:

Quality: 469.00 Length: 181  
Ratio: 3.350 Gaps: 0  
Percent Similarity: 77.348 Percent Identity: 48.066

alignment\_block:  
US-09-712-768-2 x US-08-998-416-289 ...

Align seg 1/1 to: US-08-998-416-289 from: 1 to: 856

```

333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPr 349
      ||| ::::| |::| |::| |::| |::| |::| |
4 CTGTAATACAGACTTATTTACTTCACTACATTAATTAATTCCTTATTC 53
      ::::::::::| |::| |::| |::| |::| |::| |
349 oThrGlyIleLysIlePheSerTrpIleAlaTrpMetTrpGlyGlySerV 366
      ::::::::::| |::| |::| |::| |::| |::| |
54 TACTAGTATTAAAGTATTAGTGAATTAATTAATTAATTAATTAATTAAT 103
      ::::::::::| |::| |::| |::| |::| |::| |
366 aLGIuPheLysSerPrometLeuTrpAlaPheGlyPheMetPheLeuPhe 382
      ::::::::::| |::| |::| |::| |::| |::| |
104 TAAGATTACCAACCAATATTAATTAATTAATTAATTAATTAATTAATTA 153
      ::::::::::| |::| |::| |::| |::| |::| |
383 ThrValGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAspAr 399
      ||||| |::| |::| |::| |::| |::| |::| |
154 ACTGTAAGTGTGTTACTGCTGAGTATTAATGCTAATCTATCATTAATG 203
      ::::::::::| |::| |::| |::| |::| |::| |
399 gAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMetS 416
      ::::::::::| |::| |::| |::| |::| |::| |
204 AGCATTCATGATACATTAATTAATTAATTAATTAATTAATTAATTAAT 253
      ::::::::::| |::| |::| |::| |::| |::| |
416 eLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetPro 432
      ::::::::::| |::| |::| |::| |::| |::| |
254 GTTATGCTGCTGATTTCTGATGTTGCTGCTGATTAATTAATTAATTAAT 303
      ::::::::::| |::| |::| |::| |::| |::| |
433 LysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTr 449
      ||| ::::| |::| |::| |::| |::| |::| |
304 CTGTTTAAAGTTAAATTAATGAAGAAATTAATCACAATTCATTCG 353
      ::::::::::| |::| |::| |::| |::| |::| |
449 PThrPheHeliGlyAlaAsnValThrPhePheProGluHisPheLeuG 466
      ::::::::::| |::| |::| |::| |::| |::| |
354 ATTAATTTCTAGTCTTAATTAATTTCTCTCCATGATTCCTAG 403
      ::::::::::| |::| |::| |::| |::| |::| |
466 LysArgGlnGlyMetProArgArgTyrIleAspTyrProGluAlaPheAla 482
      ||| ::::| |::| |::| |::| |::| |::| |
404 GTATTAATGATACCAAGAAATTCCTGATTAATCCGATCATTCCTA 453
      ::::::::::| |::| |::| |::| |::| |::| |
483 LeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPh 499
      ||| ||||| |::| |::| |::| |::| |::| |
454 GGTGAATTTAGTATCTTCATTTGGTTCCTATTAATCACTATATCATTT 503
      ::::::::::| |::| |::| |::| |::| |::| |
499 eLeuPheHeliValIlePheValTyrThrLeuValAlaGly 513
      ::::::::::| |::| |::| |::| |::| |::| |
504 AATGTTATCTTATATTATTATGATCAATTAATAATGTT 546
      ::::::::::| |::| |::| |::| |::| |::| |

```

seq\_name: /cgn2\_6/protdata/2/lna/6B\_COMB.seq:US-09-385-982-343

seq\_documentation\_block:

```

; Sequence 343, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639

```





; Patent No. 5741666

254 ATATAGGAATAGTATGACCAATA 276

254 ATATAGGAATAGTATGAGCAATA 276







US-09-712-768-2 x AUI75705 ..

Align seg 1/1 to: AUI75705 from: 1 to: 1054

```

210 GYMETThLeuHISLysValProLeuPheSerTrpSerIlePheIleTh 226
|||||.....:|||||.....:|||||
10 GGTATGACATAGACCGAATACCGTTATTCGTGCGGACGCTTATTTAC 59
226 r1atrrpleuIleLeuLeuAlaLeuProValLeuAlaGlyAlaIleThm 243
|||||.....:|||||.....:|||||
60 TGCACCTTTACTCCTATTAATCTTACCTGCTTGCAGAGGATATTACA 109
243 eLeuLeuThrasparGAsnPhgIlyThrTrhPhePheAsnProAlaGly 259
|||||.....:|||||.....:|||||
110 TACTCTTACCGATCGTAACTTAACTCACTTATTTTGGCCACGCGGT 159
260 GLYGlyAsnProIleLeuTyrcIleHISLeuTrpPhePheGlyHISPr 276
|||||.....:|||||.....:|||||
160 GGAGGAGACCCAGTCCTTATCAACATTATTCGTCTCGGCGACCC 209
276 oGlyValTyrcIleIleLeuProGlyPheGlyIleIleSerHISVal 293
|||||.....:|||||.....:|||||
210 AAGAGTCATATTCATATTCCTCCGCGCTTGGATTAATCTCACATCA 259
293 aISerThrPheSer...LysLysProValPheGlyTyrcLeuProMetVal 308
|||||.....:|||||.....:|||||
260 TCAGTCAGAGAGTCAGTAAAGAGCTTTGGTACTTTAGCGATATTT 309
309 TyrcIleMetValAlaIleGlyValLeuGlyPheValValTrpAlaHIS 325
|||||.....:|||||.....:|||||
310 TATGCTATATTTGGCCATTTGAGTTTGTAGGTGGCGCACCA 359
325 smETyrcThrValGlyMetSerLeuThrcIleGlnISerTyrcPheMetLeu 342
|||||.....:|||||.....:|||||
360 TATATTACAGTAGAATGATGATGATACAGACGACTTACTTACATAG 409
342 IaThrMetValIleAlaValProThrcIlyIleLysIlePheSerTrpIle 358
|||||.....:|||||.....:|||||
410 CAACATATATTTATTTGCTGCTCTACAGAAATTAATCTTATTTAGT 459
359 AlaThrMetTrpGlyGlySerValGluPheLysSerProMetLeuTrpAl 375
|||||.....:|||||.....:|||||
460 GGGAGCCCTCCATGGAACCACTAACTAAGCCATATTAATTTGGGC 509
375 aPheGlyPheMetPheLeuPheThrcValGlyValThrcIlyIleVal 392
|||||.....:|||||.....:|||||
510 GTTAGAGATTTCGTTCTTATTTACTGCGGGGAGTAAACAGAGTTGTC 559
392 euAlaGlnAlaGlyLeuAspArgAlaTyrcHISAspThrcTyrcValVal 408
|||||.....:|||||.....:|||||
560 TAGCGAATTCCTTCGATGATATTCACAGACACTTATTTATTTAGT 609
409 AlaHISpHeHISTyrcValMetSerLeuGlyAlaIlePheAlaIlePheAl 425
|||||.....:|||||.....:|||||
610 GCCCATTTCCACTACGTTCTCTCGATAGGAGCTGATTTGGAAATTTTGC 659
425 aGlyIleTyrcPheTyrcMetProLysPheSerGlyArgAlaPhe...ProG 441
|||||.....:|||||.....:|||||
660 AGCAATTTGCTCATTTGTTCTCTTTTACGTGCTCACATTAAATCTTA 709
441 LuTrpAlaAlaLysLeuHISpHeTrpThrcPhePheIleGlyAlaAsnVal 457
|||||.....:|||||.....:|||||
710 AATGG...CTAAAAATCCACTTTTATTAATTTGTTGGAGTAAACAT 756
458 ThrPhePheProGlnHISpHeLeuGlyArgGlnGlyMetProArgArgTy 474
|||||.....:|||||.....:|||||
757 ACATTTTCCCCCAACATTTCTTACGACTTAATGSAATACTCGACGATA 806
474 rILeAspTyrcProGlnAlaPheAlaLeuTrpAsnLysValSerSerTyrg 491
|||||.....:|||||.....:|||||
807 CTTGATTAATCTGATGCTACACAGCTTGAATGTGTTTCATCTATTTG 856
491 LyAlaPheLeuAlaPheAlaSerPheLeuPheIleValIlePheVal 507

```

```

|||||.....:|||||.....:|||||
857 GGTCT.....:ACCAATTTCAATTATTCCTGCTCCGAGATTTA 894
507 lTyrcThrcLeuValAlaGlyArgGlyThrArgProAsnProTrpGly 524
|||||.....:|||||.....:|||||
895 TTATCATTTGATGAGAGCTTAACCTGCTGCACGCGGTATTTATTTCTCTC 944
524 lPhe...AlaAspThrcLeuGlyTrpThrcLeuProSerProProAlaH 540
|||||.....:|||||.....:|||||
945 TATTTCTTACCACTATTCATTTGATGAGACATATTTCTCCGCGAGCTATC 994
540 lThrPheGlyThrLeuPro 546
|||||.....:|||||.....:|||||
995 ATAGTTATATGAATTCGA 1014
seq_name: gb_est1:AUI75493

```

```

seq_documentation_block:
LOCUS AUI75493 1012 bp mRNA 05-JUL-2001
DEFINITION AUI75493 Marsupinaeus japonicus adult Marsupinaeus japonicus cDNA
clone PJH742 reverse similar to Penaeus monodon cytochrome c
oxidase subunit I (AA043374), mRNA sequence.
ACCESSION AUI75493
VERSION AUI75493.1 GI:14621900
KEYWORDS EST.
SOURCE Marsupinaeus japonicus.
ORGANISM Marsupinaeus japonicus.
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
Penaeidae; Marsupinaeus.
1 (bases 1 to 1012)
Rojinakorn,J., Hirano,I., Aoki,T., Itami,T. and Takahashi,Y.
Gene expression in non-infected and viral infected hemocytes of
kuruma prawn (Penaeus japonicus)
Unpublished (2001)
CONTACT: Takashi Aoki
AQUATIC BIOSCIENCES
Tokyo University of Fisheries
Konan 4-5-7, Tokyo, Minato-ku 108-8477, Japan
Tel: 81-3-5463-0689
Fax: 81-3-5463-0690
Email: aoki@tokyo-u-fish.ac.jp.
LOCATION/Qualifiers
1..1012

```

```

FEATURES
source
1..1012
/organism="Marsupinaeus japonicus"
/db_xref="taxon:27405"
/clone="PJH742"
/clone_lib="Marsupinaeus japonicus adult"
/cell_type="hemocytes"
/dev_stage="adult"
/note="Vector: lambda ZAP II"

```

```

BASE COUNT 276 a 189 c 198 g 349 t
ORIGIN

```

```

alignment_scores:
Quality: 980.50 Length: 342
Ratio: 3.565 Gaps: 5
Percent Similarity: 80.409 Percent Identity: 54.678

```

```

alignment_block:
US-09-712-768-2 x AUI75493 ..

```

```

Align seg 1/1 to: AUI75493 from: 1 to: 1012

```

```

24 SerThrasnHISLysAspIleGlyLeuTyrcIleValAlaIleGlyVal 40
|||||.....:|||||.....:|||||
3 TCATAAACCATTAAGACATCGGACACTATATTTATTTTCGGGCGCTG 52
40 lValGlyPheIleSerValLeuPheThrcValTyrcMetArgLeuLeu 57
|||||.....:|||||.....:|||||
53 GGCAGGATATGATGCTCTCTTATGTTATTTGCTGCTGAGTTAG 102

```

```

57 etasproglyvalglntrymetcysleugllyalaargleuileala 73
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
103 GTCACACGAGA.....ACACTTATTGCA 125
74 Aspalaserglnthrcysthralaasnghlhisleutrgpsnvalmetva 90
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
126 GAT.....GACCAATCTATATATAGTAGT 151
90 lthryhiglylleulemetphephevalglylleproalaaleu 107
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
152 TACAGCTCATGCACTTTGTATATTTCTTATAGTATACCTATATATA 201
107 heglyglypheglyasntryleuemetproleuglnileglyalaproasp 123
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
202 TTGGAGATTGGAACTGGTATGCTTAAATATATAGAGACACCTGAT 251
124 Metlaphroargmetasnansleuserpheprrhephelealagl 140
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
252 ATGGCATTCACAGAAATGAATATATAGCTTTGGCTGCTGCTCTTC 301
140 ythrAlameglyvalAlaserleuphealaproglylyaspglylnl 157
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
302 T.....CTTACTCTCTCTTATACAGAGAAATAGTTGAA 336
157 euglyserglyvalglytrpvalleutyrproproleuserthrarg... 172
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
337 GAGAGATTGAGACAGATGAGCTGTACCTCTTATACAGCTAGATC 386
173 ...GluhlaglytyrsermetaspleuAlalephealavalhisleuse 188
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
387 GCACATTCAGGGGCTGTAGATTAGAAATTTTTCATTACCTTTAGC 436
188 rglAlaserSerllemetcylalaileasmetllethrthPheleua 205
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
437 AGGGGTTTCATCATTTTAGGGCCGTAATTTATACAAACGTATATA 486
205 smetargalaproglymetthrleuhslyvalproleupesertrp 221
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
487 ATATACATCTACAGGTATGATAGCCGAATACCGTATTCGTCTG 536
222 SerllephelethrAlatrpleuileleuAlaleuProvalleual 238
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
537 GCAGCTTTTACTGCACTTTTACTCTTATCTTATCTTACTGTTCTAGC 586
238 aglyAlailethrmetleuethrAsparGasnphleglythrthPhep 255
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
587 AGGAGCTATTCATCTTCTTACGATCGAATCTTAAATACCTCATTTT 636
255 heasnProAlaglyglyAsprolleuetyrglnhisleuLeutrp 271
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
637 TTGACCCAGCGGGTGAGAGACCCAGCTCTTATCAACATTATTTCTG 686
272 PhepheglyhispProgluvaltyrlelleleuProglypheglyl 288
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
687 TTCTTCGGGACCCAGAGCTTATATTTCTTCCCGCTTGGAAAT 736
288 eliseserhisvalserthrPhep...LyslysprovalPheglyt 304
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
737 AATCTCACATCATCATGCAAGAGCGGGTAATAAAGAAAGACTTTGCTA 786
304 ytleuprometvaltyrAlamevalAlaleglyvalleuglypheval 320
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
787 CTTTAGGATAATTTATGCTATATTTGGCAATTTAGGATTTGGTA 836
321 ValtrpAlahisismetlyrthrValglymetSerleuthrglnse 337
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
837 GTGGGGCTCACCATATATTACAGTAGAATGATGTGATACACAGC 886
337 rtyrPheMetleualatrhmetvallealavalprothrGlylleuys 354
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
887 TTACTTTTACATCAGCAACAATAATTTATTTGCTGCTTACAGGAATTAAT 936
354 lephesertripleAlatrhmettrp 362

```

```

937 CTTTACTTGCCTTGGACCTCCATCG 962
seq_name: gb_est1:AU167717
seq_documentation_block: 1124 bp mRNA EST 29-JAN-2001
LOCUS AU167717 1124 bp mRNA EST 29-JAN-2001
DEFINITION AU167717 O1-br-ad cDNA oryzias latipes cDNA clone br0955, mRNA
sequence.
ACCESSION AU167717
VERSION AU167717.1 GI:12589786
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Neoteleostei;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE
1 (bases 1 to 1124)
Mita,K., Ishikawa,Y. and Yamauchi,M.
AUTHORS Establishment of cDNA database of medaka, Oryzias latipes
TITLE Unpublished (2001)
JOURNAL Contact: Mita K
COMMENT Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmk@nirs.go.jp
method:uni-directional sequence direction:sequenced from T3 primer
(5' -> 3').
FEATURES
source
Location/Qualifiers
1..1124
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="br0955"
/clone_1bp="O1-br-ad cDNA"
/sex="female/male mixed"
/rissue_type="brain"
/dev_stage="adult"
BASE COUNT 292 a 284 c 210 g 338 t
ORIGIN
alignment_scores:
Quality: 946.50 Length: 344
Ratio: 3.392 Gaps: 4
Percent Similarity: 81.105 Percent Identity: 53.779
alignment_block:
US-09-712-768-2 x AU167717 ..
Align seg 1/1 to: AU167717 from: 1 to: 1124
207 ArgAlaproglymetthrleuethrleuhslyvalproleupesertrpserl 223
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2 AAACCCCGCCCATCTCTCAATACCAACCCCTTATATTTGTTGAGCCGT 51
223 ephellethrAlatrpleuileleuAlaleuProvalleualagly 240
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
52 ACTAATCACCGGAGATATTTACTCTTCTCTCCCTGCTGAGCTGAG 101
240 lalthrmetleuethrAsparGasnphleglythrthPhepheasn 256
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
102 GGATCATCTACTCTTACGACAGCCGAACCTTAATAACACTTCTCGAC 151
257 ProAlaglyglyAsprolleuetyrglnhisleuLeutrphep 273
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
152 CCGGAGGAGAGGAGCCCTATTTCTACCAACACTTTTCTGATTTCT 201
273 eglyhispProgluvaltyrlelleleuProglypheglylleuys 290
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
202 TGCCACCTGAAATTATTTCTTATTTCTCCCGGCTTGGGAATGATT 251

```

```

290  erh1zvalvalserthrpheser...LysLysProvalPheGlyThrLeu 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252  CTCATATTGCTGCTATTATTCAGGTAAAGAACCTTCGGGTACATG 301
306  PrometValTyrAlaMetValAlaIleGlyValIleuGlyPheValTyr 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302  GGAATGGATGAGCCATGATGCTATCGACCTTCAGGGTTATATGTTG 351
322  pAlaIsh1smetTyrThrValGlyMetSerLeuThrGlnGlnSerTyrP 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352  AGCTACACCATATTCACCGGTAGGATGAGCTGACACTGCAGCTTATT 401
339  hemetLeuAlaThrMetValIleAlaValIProThrGlyIleLysIlePhe 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402  TTACATCTGCACACATTAATCATTCGATTCCTAGCGGGTTAAGGTATT 451
356  SerTrrPleAlaThrMetTrrPglYserValGluPheLysSerProme 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452  AGCTGACTTGCTACCTTCATGAGAGTTCATTAATGAGAAACCCCACT 501
372  tLeuThrAlaPheGlyPheMetPheLeuPheThrValGlyValIThrG 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502  ATTATGGGCCCCCTGGTTTATTATTTCTCTTTACACTGGGGGCTTACTG 551
389  lYlIleValIleuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyr 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
552  GCATGTGCTTAGCCCAATCATCTTAGATATTATCTGCATGACACTTAC 601
406  TyrValValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAl 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
602  TATGTAGTAGCTCACTTCACATATGCTCTATCAATAGGGGCGCTATCGC 651
422  allePheAlaGlyIleTyrPheTyrMetProLysPheSerGlyArgAlaP 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652  TATTATGGGGGCGATTCGTTCACTGATTCCTCTATTTCCGGCTACTT 701
439  hePrrGlyTrrPAlaAlaLysLeuHisPheTrrPhePheIleGlyAla 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
702  TACACGGAACATGAGAACGATCCACTTCGAGATATTTGTGGGGGTA 751
456  AsnValIThrPhePhePrrGlnHisPheLeuGlyArgGlnGlyMetPrrAr 472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
752  AATTATACCTTCTTCCTCAACACTTCTTAGCGCTAGCGGAATACCAAG 801
472  gArGlyTrrLeaSprTyrProGlnAlaPheAlaLeuTrrPasnLysValSer 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
802  AGCATACCTGCTACCTACCAAGCCCTATACACTAGAAACACTATTTCAT 851
489  erTyrGlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleValIle 505
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
852  CCTCGGCTCCCTTATCTCAGCTTGTGCTGATATATTCCTATTTCAT 901
506  PheValTyrThrLeuValAlaGlyArgArgGluThrArgProAsnProTr 522
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
902  CTATG....AGAACGCTTGACAGCTAACGTGAAGTTCTG..... 936
522  pGlyGluPheAlaSprThrLeuGluTrrThr.....L 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
937  .....TCGGTGACCTGACTGCGCCCAACGTAAGATGAC 970
533  euPrrSerProPrrAlaHisThrPheGlu 543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
971  TCCCGATGCCCTCCCTTATACACTTTTGAA 1002
seq_name: gb_est1:BE468621
seq_documentation_block:
LOCUS BE468621 958 bp mRNA EST 28-JUL-2000
DEFINITION IPHdK00496 Head kidney cDNA library Ictalurus punctatus cDNA
similar to Cytochrome c oxidase subunit I, mRNA sequence.
ACCESSION BE468621
VERSION BE468621.1 GI:9559112
KEYWORDS EST.
```

```

SOURCE channel catfish.
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Siluriformes; Ictaluridae; Ictalurus.
REFERENCE 1 (bases 1 to 958)
AUTHORS Cao,D., Kocabas,A., Ju,Z., Karsli,A., Li,P., Patterson,A. and Liu
,Z.J.
TITLE Transcriptome analysis of channel catfish (Ictalurus punctatus):
Gene cataloguing and profiling from the head kidney
JOURNAL Unpublished (2000)
COMMENT The Fish Molecular Genetics and Biotechnology Laboratory,
Contact: Liu ZJ
Department of Fisheries and Allied Aquacultures and Program of Cell
and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: M3 Reverse
FEATURES
Location/Qualifiers
source 1..958
BASE COUNT 237 a 259 c 180 g 282 t
ORIGIN
alignment_scores:
Quality: 905.50 Length: 341
Ratio: 3.269 Gaps: 6
Percent Similarity: 81.232 Percent Identity: 53.372
alignment_block:
us-09-712-768-2 x BE468621 ..
Align seg 1/1 to: BE468621 from: 1 to: 958
19 ThArGTrPheMetSerThrAsnHisLysAspIleGlyLeuTyrIle 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 ACGGCGGATTTTTCACACCAACATTAAGATATGCGACCCCTTAAC 51
35 uValAlaIleGlyValAlaGlyPheIleSerValLeuPheThrValTyrM 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 TGTATTGGTGCCGACCGGAAATAGGGGTACCGCCCTTAAGCCTGTTA 101
52 eTarGLeuGluLeuMetAspProGlyValGlnTyrMetCysLeuGluGly 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 TCGGGGCGGATTTAGCCCAACCCGGCGC..... 130
69 AlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnGlnHisIle 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 .....CTCTAGGCGAT.....GACCAAT 150
85 uTrrPasnValMetValThrTyrHisGlyIleLeuMetMetPhePheValG 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 TTACAAATGTTATTGTTACTGCTCAGCGCTTTGTAATAATTTCTTTATAG 200
102 lYlIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPrrLeuGln 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 TATATACCAATTATGATTTGAGGGTTTGCAACTGCTTGTTCCTTAATG 250
119 lIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeuSerPheTr 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 ATCGAGACGCGCAGATATGCGCTTCCTCGAATGACAACATTAAGCTTCG 300
135 pLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAlaPrrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 GCTTCCT.....GCCCTTCCTTCCTTACTTCGCTCGCTCGCT 337
```





```

225 leThrAlaTrpLeuIleLeuAlaLeuProValIleuAlaIle 241
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 TTACGACATTTACTCCATTAATCTTACCTGCTACACAGACTATT 534
242 ThrMetLeuLeuThrAspArgAsnPhelGlyThrPhePheAsnProAl 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
535 ACAAAATCTTCTACGATCGTAACTTAATCTTACTTATTTTGACCAGC 584
258 acGlyGlyAspProIleLeuTyGlnHisIleLeuTrpPhePheGlyH 275
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
585 GGGTGAGAGACCAAGTCCTTATCAACATTATCTGATCTTCCGCGC 634
275 iAProGluValIleIleIleIleLeuProGlyPheGlyIleIleSerHis 291
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
635 ACCCAGAGCTATATCTCAATCTTCCCGCTTGGAATATCTCACAC 684
292 ValValSer...ThrPheSerLysLysProValPheGlyTyTrLeuProme 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
685 ATCATCATGCAAGAGCGGGTAAAGAGAGCTTTGGTACTTATGGGAT 734
307 tValTyAlaMetValAlaIleGlyValLeuGlyPheValValTrpAlaH 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
735 AATTATAGCATATTGGCATTGGAGTTTATGAGATTGTAGTGAGCTG 784
324 IShiMetLysThrValGlyMetSerLeuThrGlnGlnSerTyPheMet 340
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
785 ACCATATATTTCATAGTAATGATGTTGATACACGAGCTTACTTTACA 834
341 LeuAlaThrMetValIleAlaValProThrGlyIleLysIlePheSerTr 357
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
835 TCAGCAACAATTAATTATGCTGCTCTACAGAAATTAATC..... 876
357 PLeuAlaThrMetTrpGlyLysSerValGluPheLysSerPromeLeuTr 374
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
877 .TTAGTGAAGTGAAGCTCAAGACCACTAAACATATACCATCATTAAT 925
374 tPAlaPheGlyPheMetPheLeuPheThrValGlyGlyValThrGlyIle 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
926 GGGCGTTAGGATCGGTTCTATTAATGATGATGAT.....ACAGAGATG 969
391 ValLeuAlaGlnAlaGlyLeuAspArgAlaTyTrHisAspThr 404
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
970 .....TCTAGCGATCTGATGATCATATCTACGACACT 1002

seq_name: gb_est1:BE574185

seq_documentation_block:
LOCUS BE574185 961 bp mRNA EST 15-AUG-2000
DEFINITION IPBr01227 Brain cDNA library Ictalurus punctatus cDNA 5' similar
to Mitochondrion tRNA(Ser) (UCN), tRNA(Asp) and cytochrome c
oxidase subunits I and II genes, mRNA sequence.
ACCESSION BE574185
VERSION BE574185.1 GI:9824230
SOURCE EST.
ORGANISM channel catfish.
Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Siluriformes; Ictaluridae; Ictalurus.
REFERENCE 1 (bases 1 to 961)
AUTHORS Ju,Z., Karsl,A., Kocabas,A., Patterson,A., Li,P., Cao,D., Dunham,R,
and Liu,Z.
TITLE Transcription analysis of channel catfish: I. genes and expression
profiles from the brain
JOURNAL Unpublished (2000)
COMMENT Contact: Liu Z
The fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell
and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054

```

```

FEATURES
    source          .
                    /organism="Ictalurus punctatus"
                    /db_xref="taxon:7998"
                    /clone_lib="Brain cDNA library"
                    /note="Organ: Brain; Vector: pSport1; Site_1: NotI;
                    Site_2: SalI"
BASE COUNT      247 a      266 c      166 g      282 t
ORIGIN
alignment_scores:
    Quality:      890.50      Length:      301
    Ratio:        3.425      Gaps:      1
    Percent Similarity: 86.379      Percent Identity: 55.814
alignment_block:
US-09-712-768-2 x BE574185
Align seg 1/1 to: BE574185 from: 1 to: 961
207 ArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSerI 223
    :: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1 AAGCCCCCGCATCTCATCATATCAACACCCCTATTGCTGAGCGCT 50
223 ePheIleThrAlaTrpLeuIleLeuAlaLeuProValIleuAlaGlyA 240
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
51 CCTATTACAGCCGCTCCTCACTACTATCCCTCCAGTTTATGACCGCTG 100
240 lAlaIleThrMetLeuThrAspArgAsnPhelGlyThrTrpPhePheAsn 256
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
101 GTATCAACAATTAATTAACAGACCAAACTTAATTAATTAATTAATTAAT 150
257 ProAlaGlyGlyAspProIleLeuTyGlnHisIleLeuTrpPheP 273
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
151 CCCGACGAGAG. GCGACCCCATCTTTACCACACCTTTCTGATTTCT 199
273 eGlyHisProGluValIleIleIleLeuProGlyPheGlyIleIles 290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 TGGGACACCCAGAAATTAATTAATTTGACAGGCTTGGCATATCT 249
290 eHisValIleSerThrPheSer...LysLysProValPheGlyTyTrLeu 305
    ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
250 CCATATTGTTGCTCACTACGACGCAAAAAGAACCATTCGGCTATATG 299
306 PromeValTyAlaMetValAlaIleGlyValLeuGlyPheValTr 322
    ::||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
300 GGATTAAGTGAAGTATATAGCATTTGGCTTCTAGGCTTTATGTATG 349
322 pAlaHisMetLysThrValGlyMetSerLeuThrGlnGlnSerTyTr 339
    ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
350 AGCATCATCATATTAATCACTAGAGATAGAGCTGACATCGAGCATAT 399
339 hemeLeuAlaThrMetValIleAlaValProThrGlyIleLysIlePhe 355
    ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
400 TTACATGTGACACAAATTAATTAATCCCACTGGGGTCAAGATAT 449
356 SerTrpIleAlaThrMetTrpGlyLysSerValGluPheLysSerProme 372
    ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
450 AGGTGACTTGCTACCTACACAGGAGGCTCATTAATAGAAATCCCT 499
372 tLeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyValThrG 389
    ::||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
500 GCTCTGAGCCCTAGGTTTCATCTTCTTACTGTTGGGGCTTAATG 549
389 lTleValIleuAlaGlnAlaGlyLeuAspArgAlaTyTrHisAspThrTy 405
    ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
550 GCATTGACTAGCCAACTCATCCCTAGACATGTTGTTACATGACCTTAC 599
406 TyValValAlaHisPheHisTyValMetSerLeuGlyAlaIlePheAl 422

```

```

|||||
600 TATGTATACCCATTCATATGTGTATCAATGAGGGCTGATTTGC 649
422 attlephalaglylethyrpheyrmetyrolyspheSerGlyArgAlar 439
650 TATTATGGGGCCCTTCCTGCTACGTGATCCCTCTTACAGATACACAA 699
439 heproGluTPAlaAlaLysLeuHisPheThrPhePheleIleGlyAla 455
700 TACAGACACACTGACAAATAATTCACCTCGGAACTATATTGTGGGGGTA 749
456 AsnValThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProAr 472
750 AACTCACCCTTCTCCCAACACTTCTAGCCCTCCGTCGATACACAG 799
472 gArgTyrIleAspTyrProGlnAlaPheAlaLeuTPAsnLysValSerS 489
800 ACGATCTACAGACTACCCGATGCTACTCATATGAACATGATCTCTT 849
489 eRTYrGlyAlaPheLeuAlaPheAlaSerPheLeuPheleIleValIle 505
850 CAATGGGCTCCTGCTGCTCCCTAGTACAGCTTGTATTCCTGATATT 899
506 Phe 506
900 TAT 902

seq_name: gb_est2:BG784814
seq_documentation_block:
LOCUS BG784814 784 bp mRNA EST 20-MAY-2001
DEFINITION SEAMC004771 sea urchin primary mesenchyme cell cDNA library
strongylocentrotus purpuratus cDNA clone PC_0023_A1_H09_MR 5', mRNA
sequence.
ACCESSION BG784814 GI:14155827
VERSION BG784814.1 GI:14155827
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoda;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 784)
AUTHORS Zhu,X., Mahaltras,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
TITLE A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
JOURNAL Development 128 (2001) In press
COMMENT Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.

FEATURES
source
1..784
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="PC_0023_A1_H09_MR"
/clone_lib="Sea urchin primary mesenchyme cell cDNA
library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
at priming from poly A+ RNA, directionally cloned"

BASE COUNT 184 a 184 c 157 g 256 t 3 others
ORIGIN
alignment_scores:
Quality: 875.50 Length: 260
Ratio: 3.908 Gaps: 1

```

Percent Similarity: 86.154 Percent Identity: 61.154

alignment\_block:

US-09-712-768-2 x BG784814 ..

Align seg 1/1 to: BG784814 from: 1 to: 784

```

234 LeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsn 250
5 ATTCCCGGGTCGACCCAGCCGCTCCGCAATGCTTTCACAGATCGTAAT 54
250 eGlyThrThrPhePheAsnProAlaGlyGlyGlyAspProIleLeuTyrG 267
55 AACACAAACTTCTTGAACCTCGACAGAGGGGAGATCAATTCGTTTC 104
267 IHisIleLeuThrPhePheGlyHisProGlnValTyrIleIleIleLeu 283
105 AACACCATTCGTTGTTTGGACACCCCGAGGTGTATATTCCTTATTC 154
284 ProGlyPheGlyIleIleSerHisValValSerThrPheSer...LysLy 299
155 CCGGATTTGGTATGATCTACACGTTAATAGCTCACTCTGTAACG 204
205 AGACCTTCGGATACCTGGGAGGATGTTATGCCATGATTCGATAGAG 254
299 sProValPheGlyTyrLeuProMetValTyrAlaMetValAlaIleGly 316
316 AlLeuGlyPheValValTyrPalaHisHisMetTyrThrValGlyMetSer 332
255 TTTTAGGATTCCTTGTCTGGGCCACCAATGTCTTACAGTGAATGGAT 304
333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPr 349
305 GTTGATACACGACACTACTCTACCTCCGCGACAGATGATATTGCGTTCC 354
349 OThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGlyGlySerV 366
355 AACAGGATTAAGGTTTCAGATGAATGCAACGCTCCAGAGGCTAATC 404
366 AlGluPheLysSerProMetLeuTyrPalaPheGlyPheMetPheLeu 382
405 TACAAAGAAACTCCCTTATATATAGACCTTGGAATTTGTATTATTATT 454
383 ThrValGlyGlyValThrGlyLeuValLeuAlaGlnAlaGlyLeuAspAr 399
455 ACATTAGAGAGACTCACAGTATGTGCTTGCCAAATCCCATGAGCGT 504
399 gAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMetS 416
505 TGTTCTCATGATACCTACTACGTGAGTGCACCTTTCACATGAGCTTCT 554
416 eLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetPro 432
555 CAATGGGGCTGATTTGCAATCTTCGCTGGTTTCACACAGGTTTCC 604
433 LysPheSerGlyArgAlaPheProGluTPAlaAlaLysLeuHisPheTr 449
605 CTTCTTCGTTATAGACCTACACCATATAGAGAAAGGTTCACTCTT 654
449 pThrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuG 466
655 CATTAAGTTTGTGGAGTCACCTTAACCTTTTTCCTCAACCTTCTTA 704
466 LysArgGlnGlyMetProArgArgTyrIleAspTyrProGlnAlaPheAla 482
705 GCTACCGCGGATGCCACGACGATCTCAGACTATCCAGAGCCATATACA 754
483 LeuTPAsnLysValSerSerTyrGlyAla 492
755 CTTTGAATATCATCTNCATTCATTTGATACA 784

seq_name: gb_est1:AJ395912

```

```

seq_documentation_block:
LOCUS      AJ395912      805 bp      mRNA      25-JAN-2001
DEFINITION AJ395912 dkt426 Gallus gallus cDNA clone 24j17c1, mRNA sequence.
ACCESSION  AJ395912
VERSION     AJ395912.1 GI:7127314
KEYWORDS    EST.
SOURCE      chicken.
ORGANISM    Gallus gallus
             Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
             Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
             Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 805)
AUTHORS     Adirakhanov, I., Lodygin, D., Geroth, P., Arakawa, H., Law, A., Plachy
             J., Korn, B. and Buerstedde, J.M.
TITLE       A large database of chicken bursal ESTs as a resource for the
             analysis of vertebrate gene function
JOURNAL     Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE     20568495
COMMENT     Contact: Buerstedde JM
             Cellular Immunology
             Heinrich-Pette-Institute
             Martinistr. 52, 20251 Hamburg, Germany
             Email: URL: http://genetics.npl.uni-hamburg.de/dt4dest.html.
             Location/Qualifiers
FEATURES
             source
             1. 805
                /organism="Gallus gallus"
                /strain="CB"
                /db_xref="taxon:9031"
                /clone="24j17c1"
                /clone_lib="dkt426"
                /tissue_type="Bursa of Fabricius"
BASE COUNT  214 a      252 c      136 g      203 t
ORIGIN
alignment_scores:
            Quality: 872.50      Length: 266
            Ratio: 3.729      Gaps: 2
            Percent Similarity: 87.970      Percent Identity: 61.654
alignment_block:
US-09-712-768-2 x AJ395912 ..
Align seg 1/1 to: AJ395912 from: 1 to: 805
156 GlnLeuGlySerGlyValGlyTyrPValLeuTyrProProlen.....Se 170
    :::::::::::::::::::::
    7 GAAGCTGGGGCGCGCAGAGATGACAGTTTACCCCTTTAGCCGGCAA 56
170 rThrArgGuaIaGlyTyrSerMetAspLeuAlaIlePheAlaValHisL 187
    ::::::::::::::::::::
57 CCTACCCACGCTGGCGCATGATGACCTAGCCATCTTTTCATATACACT 106
187 euSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIleThrThrPhe 203
    ||:::::::::::::::::::
107 TAGCAGGGGTTCCTCCATTTAGAGAGCCATCAACTTATCATCTACATC 156
204 LeuAsnMetArgAlaProGlyMetThrLeuHisGlyValProlenPheSe 220
    :::::::::::::::::::::
157 ATCAACATTAACCCCGCAGCTGTCAATAACCAACACCCCTATTGT 206
220 rTyrSerIlePheIleThrAlaTyrPLeuIleLeuAlaLeuProValL 237
    ||:::::::::::::::::::
207 ATGATCCGCTCATCTACTAGCCATCTACTCTCTCTCTTACCCGCC 256
237 euAlaGlyAlaIleThrMetLeuThrAspArgAspPheGlyThrThr 253
    ||:::::::::::::::::::
257 TAGCAGCTGGGATTACATCTACTTACGAGCCGACACTTAACACGACA 306
254 PhePheAspProAlaGlyGlyAspProIleLeuTyrGlnHisIleLe 270
    ||:::::::::::::::::::
307 TTCTTCGAGCCAGCTGGAGAGAGAGACCAATCTATCAACACACCTATT 356

```

```

270 uTrrPhePheGlyHisProGluValTyrIleIleIleLeuProGlyPheS 287
    ::::::::::::::::::::
357 CTGATCTCTGGTCACCCCGAAGTTTACATCTCATCTCCAGGTTGG 406
287 lylIleIleSerHisValValSerThrPheSer...LysLysProValPhe 302
    ||:::::::::::::::::::
407 GAATTAATTTCCACGTAGTACTATGACAGAAAAAAGAACCATTC 456
303 GlyTyrLeuProMetValTyrAlaMetValAlaIleGlyValLeuGlyP 319
    ||:::::::::::::::::::
457 GGATACATAGGAATAGTCTGACCATATCTCATCTCGATTGGCTT 506
319 eValValTrrPAlaHisHisMetTyrThrValGlyMetSerLeuThrG 336
    ||:::::::::::::::::::
507 CATGTGTGAGCCACCATATATTCACATCGGAGATGAGATGACACCC 556
336 lnsertTrrPheMetLeuAlaThrMetValIleAlaValProThrGly 352
    ::::::::::::::::::::
557 GAGCTTACTTTACATCAGCCACATAATATCATCCGCACTCCAGTATT 606
353 LysIlePheSerTrrPLeaIaThrMetTrrPglySerValGluPhe 369
    ||:::::::::::::::::::
607 AAAGCTTCAGCTGATGACACACCTGACAGGAGAACATTAATGAGA 656
369 sSerProMetLeuTrrPAlaPheGlyPheMetPheLeuPheThrValG 386
    ||:::::::::::::::::::
657 CCCCCCTATGCTATGACCCCTAGAGCTCATCTCTCTTACATGAGAG 706
386 lYValTrrGlyIleValLeuAlaGlnAlaGlyLeuAspArgAlaTrrH 402
    ||:::::::::::::::::::
707 GCCCTAACGGGAAATGCTCTGCTAACATCATCATGATATATGCTCAT 756
403 AspThrTrrTrrValValAlaHisPheHisTrrValMetSerLeuGly 418
    ||:::::::::::::::::::
757 GACACCTACTATGATGATGCGCCACTTTCACTATGCTCTCAATGGGG 804
seq_name: gb_est2:BG783541
seq_documentation_block:
LOCUS      BG783541      822 bp      mRNA      20-MAY-2001
DEFINITION SEUIMC003498 Sea urchin primary mesenchyme cell cDNA library
             Strongylocentrotus purpuratus cDNA clone PC_0018_B1_E11_MR 5', mRNA
             sequence.
ACCESSION  BG783541
VERSION     BG783541.1 GI:14154554
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
ORGANISM    Strongylocentrotus purpuratus
             Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
             Echinoidea; Echinoidea; Echinacea; Echinoidea;
             Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1 (bases 1 to 822)
AUTHORS     Zhu, X., Mahaitas, G., Illies, M.R., Cameron, R.A., Davidson, E.H. and
             Ettensohn, C.A.
TITLE       A large scale analysis of mRNAs expressed by primary mesenchyme
             cells of the sea urchin embryo
JOURNAL     Development 128 (2001) In press
COMMENT     Contact: Ettensohn CA
             Dept. Biol. Sci.
             Carnegie Mellon University
             4400 Fifth Avenue, Pittsburgh, PA 15213, USA
             Tel: +1 412 268 5849
             Email: ettensohn@andrew.cmu.edu.
FEATURES
             source
             1. 822
                /organism="Strongylocentrotus purpuratus"
                /db_xref="taxon:7668"
                /clone="PC_0018_B1_E11_MR"
                /clone_lib="Sea urchin primary mesenchyme cell cDNA
                library"
                /tissue_type="embryo"
                /cell_type="primary mesenchyme cells"
                /lab_host="E.coli"

```

/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; oligo  
at priming from poly A+ RNA, directionally cloned"  
BASE COUNT 189 a 203 c 166 g 263 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 864.50 Length: 269  
Ratio: 3.726 Gaps: 3  
Percent Similarity: 86.245 Percent Identity: 61.338

alignment\_block:  
US-09-712-768-2 x BG783541 ..

Align seg 1/1 to: BG783541 from: 1 to: 822

```

151 ProGlyGlyAspGlyGlnLeuGlySerGly.....ValGlyTr 163
      |||:||||| ||| ||| |||
4 CCGCAGAGTACCGGTCGGAATTCGCCGACGCCGCGCTCGGCTCG 53
      |||:||||| ||| ||| ||| ||| ||| ||| ||| |||
163 pValLeuTyProProLeuSerThrArg.....GluAlaGlyTyrSerM 178
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 AACATATACCTACCTCTCTCTCTACTAAATAGCACACGCCGAGGCTCG 103
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 etAspLeuAlaLeuPheAlaValHisLeuSerGlyAlaSerSerMet 194
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 TTGATTAGCAATCTCTCCCTCACCTCGCGGCTCTCCATCTTG 153
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 GlyAlaLeuHisMetIleThrThrPheLeuHisMetArgAlaProGly 211
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 GCGCAATTAATTAATTAACACAACTTAATTAATTAATTAATTAAT 203
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
211 tThrLeuHisValProLeuPheSerThrPheHisValLeuThrAla 228
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 GTCTTTGATCGTCTCTCTTATTCGTGATCGTCTGTCACCTCTCT 253
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 rPheLeuLeuLeuAlaLeuProValLeuAlaGlyAlaIleThrMet 244
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 TCTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 303
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 LeuThrAspArgAsnPheGlyThrThrPhePheAsnProAlaGly 261
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 CTCACAACTGTAATAATTAACACAACTTCTTGACCTCGACGAGG 353
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
261 yAspProIleLeuTyGlnHisIleLeuTrpPhePheGlyHisProG 278
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 AGATCCAAATCTCTGTTCAACACATCTCTGTTTGGACACCCGAG 403
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 aTyrlIleIleIleLeuProGlyPheGlyIleIleSerHisVal 294
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 TGTATATTCTTATCTTACCGGATTTGTATGATCTCACAGTTATAG 453
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 ThrPheSer...LysLysProValPheGlyTyrlLeuProMetVal 310
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 CACTACTCTGTGTAAGCGAGACCTTTTCGAACTGGGAGTGTATG 503
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
310 aMetValAlaIleGlyValLeuGlyPheValAlaTrpAlaHisMet 327
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
504 CATGATTCGATAGAGAGTTTAGATTCCTTGTCTGGCCACCATATG 553
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 yThrValGlyMetSerLeuThrGlnIleSerTyrlPheMetLeuAla 343
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
554 TTACAGTAGGAATGATGTATACACCATATCTCACTGCCGACG 603
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 MetValIleAlaValProThrGlyIleLysIlePheSerTrpIleAl 360
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
604 ATGATTAATTCCTCTCCACACGAGATTAANGTTTTCAGATGAATG 653
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 rMetTrpGlyGlySerValGlnPheLysSerProMetLeuTrpAlaP 377
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
654 GCTCCAGAGGCTGTAATCTACATGAGAAACCTCTTATTTAGCCCT 703
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377 yPheMetPheLeuPheThrValGlyGlyValThrGlyIleValLeu 393

```

```

|||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
704 GATTGTATTTTATTATTACATTAGGAGACCTCACAGTATGCTTGC 753
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 GluAlaGlyLeuAspArgAlaTyrlHisAspThrTyrlValAlaHis 410
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
754 AATTCCTCATTTGAGCTTGCTTCATGATACCTACTACGTGTAGTCA 803
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 sPheHis 412
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 CTTTCAC 810

```

seq\_name: gb\_est1:AU183416

seq\_documentation\_block:

LOCUS AU183416 1235 bp mRNA EST 01-MAY-2001

DEFINITION AU183416 Cyprinus carpio head kidney stimulated by  
lipo-polysaccharide and concanavalin-A Cyprinus carpio cDNA clone

ACCESSION AU183416 GI:13919157

VERSION AU183416

KEYWORDS EST.

SOURCE common carp.  
ORGANISM Cyprinus carpio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.

REFERENCE 1 (bases 1 to 1235)  
Sakai, M. and Sazan, R.  
Analysis of expressed sequence tags from a Common carp, Cyprinus

JOURNAL Unpublished (2001)  
CONTACT: Masahiro Sakai  
Miyazaki University  
Faculty of Agriculture  
1-1 nishi gakuenkibanadai, Miyazaki, Miyazaki 889-2192, Japan  
Email: m.sakai@cc.miyazaki-u.ac.jp.

FEATURES  
source

1..1235  
/organism="Cyprinus carpio"  
/db\_xref="taxon:7962"  
/clone\_lib="Cyprinus carpio head kidney stimulated by  
lipo-polysaccharide and concanavalin-A"  
/tissue\_type="head kidney stimulated by  
lipo-polysaccharide and concanavalin-A"

BASE COUNT 369 a 301 c 213 g 352 t  
ORIGIN

alignment\_scores:  
Quality: 864.00 Length: 306  
Ratio: 3.585 Gaps: 5  
Percent Similarity: 76.756 Percent Identity: 55.229

alignment\_block:  
US-09-712-768-2 x AU183416 ..

Align seg 1/1 to: AU183416 from: 1 to: 1235

```

249 AsnPheGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 265
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 AACCTTATACCACTTCTTGACCCGAGGAGGAGAGACCAATCT 279
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 uTyrlGlnHisIleLeuTrpPhePheGlyHisProGlyValTyrl 282
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 TTATCAACACTTATTCGATTTCTCGGCCACCCAGAAAGTTATATTC 329
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 LeuProGlyPheGlyIleIleSerHisValAlaSerThrPheSer... 297
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
330 TCTCTCCAGAGTTGCGTATCATCTCCATGTTGTAGCCCTATTTACGC 379
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 LysLysProValPheGlyTyrlLeuProMetValTyrlaMetValAla 314
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

380 AAAAAAGACCGTTCGGTACATAGCAATGCTGAGCCATGATGCCAT 429  
 314 eGlyValLeuGlyPheValIleValTrpAlaHisMetYrThrValGlyM 331  
 430 CGGTCTACAGGCTTATGTGTGAGCTCACCACATGTTCACTGTGGAA 479  
 331 eSerLeuThrGlnGlnSerYrPheMetLeuAlaThrMetValIleAla 347  
 480 TAGAGGTACACACTCGAGCATATTTACATCCGCAACAATATATTCGA 529  
 348 ValProThGlyIleYsIlePheSerTrpIleAlaThrMetTrpGlyG 364  
 530 ATTCACACAGGCTGTAAGATATTAGCGATTAGCACACACTCCAGGAG 579  
 364 ySerValGluPheYsSerProMetLeuTrpAlaPheGlyPheMetPhe 381  
 580 ATCCATTAAATGAGAAACCAATCTAGTGGCTTAGGGTTTATTTTCC 629  
 381 ePheThrValGlyValIleValThrGlyIleValLeuAlaGlnAlaGly 397  
 630 TATTCACAGTAGGGGAGCTCAGAGGATGCTCTATCCCAATTCATCACC 679  
 398 AspArgAlaTyHisAspThrYrTyTrpValValAlaHisPheHisTyVa 414  
 680 GATATTGCTCTCAGCAGCACTTACTAGTAGAGCAGCACTTCCACTATGT 729  
 414 lMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleYrPheTyM 431  
 730 ACTATTCATAGGTGCTGTATTTGCTATATATGACACCTTTGGCATTTAT 779  
 431 eTrpYsPheSerGlyArgAlaPheProGlyTrpAlaAlaYsLeuHis 447  
 780 TCCCCCTTAACTGATATCTCTCCAGCAGCACTGAACAAATATCCAC 829  
 448 PheTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPh 464  
 830 TTTCGAGTATGTATTGAGTAACTCAGCTTCTTCCCAACACACTT 879  
 464 eLeuGlyArgGlnGlyMetProArgTrpIleAspYrProGlyAlaP 481  
 880 CCTAGGCTTAGCGGATATGCCAGCGCATCTCCGACATATCCAGATGCT 929  
 481 hAlaLeuTrpAsnYsValSerSerYrGlyAlaPheLeuAlaPheAla 497  
 930 ACGCCCTGGAATATACAGTATCATCATGATGATCCCTAATTTCC..... 973  
 498 SerPheLeuPhePheIleValIlePheValYrThrLeu.....Va 511  
 974 .....CTAGTAGCAGTAAATCATATTCTATTATCTATGAGAGCCCTT 1017  
 511 lAlaGlyArgTrpGlnTrpArgProAsnProTrpGlyIlePheAlaAsp 528  
 1018 CGCCGCTTAACAGAGAACTA.....T 1040  
 528 hLeuGlyTrpThrLeuProSer.....ProPro 537  
 1041 CTGTAGACTAAGACCAACAATGTAGATGACTTCAAGGCTGCCCCCT 1090  
 538 ProAlaHisThrPheGly 543  
 1091 CCTTATCACACATACGAG 1108  
 seq\_name: gb\_est2:BG784767  
 seq\_documentation\_block: 787 bp mRNA EST 20-MAY-2001  
 LOCUS BG784767  
 DEFINITION SPAMUC004724 Sea urchin primary mesenchyme cell cDNA library,  
 Strongylocentrotus purpuratus cDNA clone PC\_0023\_A1\_D09\_MR 5, mRNA  
 sequence.  
 ACCESSION BG784767  
 VERSION BG784767.1 GI:14155780  
 KEYWORDS EST.  
 SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Echinoidae; Echinacea; Echinoida;  
 Strongylocentrotidae; Strongylocentrotus.  
 REFERENCE 1 (bases 1 to 787)  
 AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and  
 Etlenson,C.A.  
 TITLE A large scale analysis of mRNAs expressed by primary mesenchyme  
 cells of the sea urchin embryo  
 JOURNAL Development 128 (2001) In press  
 COMMENT Contact: Etlenson CA  
 Dept. Biol. Sci.  
 Carnegie Mellon University  
 4400 Fifth Avenue, Pittsburgh, PA 15213, USA  
 Tel: +1 412 268 5849  
 Email: etlenson@andrew.cmu.edu.  
 FEATURES  
 source  
 1. 787  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="PC\_0023\_A1\_D09\_MR"  
 /clone\_lib="Sea urchin primary mesenchyme cell cDNA  
 library"  
 /tissue\_type="embryo"  
 /cell\_type="primary mesenchyme cells"  
 /note="vector: pSPORT1; site\_1: NotI; site\_2: SalI; oligo  
 dt priming from poly A+ RNA, directionally cloned"  
 BASE COUNT 184 a 186 c 162 g 253 t 2 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 863.50 Length: 248  
 Ratio: 3.998 Gaps: 1  
 Percent Similarity: 87.097 Percent Identity: 62.500  
 alignment\_block:  
 US-09-712-768-2 x BG784767 ..  
 Align seg 1/1 to: BG784767 from: 1 to: 787  
 246 ThrAspArgAsnPheGlyThrThrPheAsnProAlaGlyGlyAs 262  
 41 ACAATGCTAAATAAACACACACTTTTGCACCTCCAGAGGGGAGCA 90  
 262 pProIleuTrpGlnHisIleLeuTrpPhePheGlyHisProGlyVal 279  
 91 TCCAAATCTGTTTCACACCACTATCTGTGTTTGGACACCCGAGGT 140  
 279 yTrIleIleIleuProGlyPheGlyIleIleSerHisValIleSerThr 295  
 141 ATATTCTTATCTTACCGGATTTGGTATGATCTCACAGCTTATAGCTC 190  
 296 PheSer...LysLysProValPheGlyTrpLeuProMetValTyrala 311  
 191 TACTCTGATAGCAGAGCGCTTGGATACCTGGGATGGTTATGCGAT 240  
 311 tValAlaIleGlyValLeuGlyPheValIleTrpAlaHisMetYrT 328  
 241 GATTGCAATAGGAGTTTATGATCTTGTCTGTGGCCACCAATATGTTTA 290  
 328 hValGlyMetSerLeuThrGlnGlnSerYrPheMetLeuAlaThrMet 344  
 291 CAGTAGGAATGATGTGATACAGACGATATCTTCACTGCCGCCAGCAT 340  
 345 ValIleAlaValProThrGlyIleYsIlePheSerTrpIleAlaThrMe 361  
 341 ATATTGCTGTTCCACAGCATTAAGTTTTCAGATGAATGCAACGCT 390  
 361 tTrpGlyGlySerValGluPheYsSerProMetLeuTrpAlaPheGly 378  
 391 CCAAGGCTTAATCTTACATGAGAACTCCCTTATATATGAGCCTTGAG 440

```

378 hemetPheLeuPheThrValGlyValThrGlyIleValLeuAlaGln 394
|||||
441 TTGATTTTATTATTACATTTAGAGAGACTCAGCATATGTTCTGCCAAT 490
|||||
395 AlAGLYLeuAspArgAlaTyrHisAspThrTyrTyrValValAlaHisph 411
|||||
491 TCCTCCATGACGGTTGTTTCATGATACCTACTAGCTGAGCTGACCTCACT 540
|||||
411 eHisTyrValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyTyr 428
|||||
541 TCACCTAGCTTCTTCATATGGGGCTGATTTGGCAATCTTGGCTGGTTCGA 590
|||||
428 YrPheTyrMetProLysPheSerGlyYarGlnPheProGluTyrPAlaIle 444
|||||
591 CTCACGTGTTCCCTCTCTCTGCTGATTATAGCTACACCCATATAGAGA 640
|||||
445 LysLeuHisPheThrPThrPhePheIleGlyAlaAsnValThrPhePhe 461
|||||
641 AAGGTTCACTTCTTCATATGTTGTTGAGTCACTTACCTTTTTC 690
|||||
461 oGlnHisPheLeuGlyYarGlnGlyMetProArgArgTyrIleAspTyr 478
|||||
691 TCACACTTCTTANGTCTAGCCGGAGTCCAGCAGGCTACTCAGACTATC 740
|||||
478 roGluAlaPheAlaLeuTyrPAsnLysValSerSerTyrGlyAla 492
|||||
741 CAGACGCTATACACTTGGAAATCTATCTGCTGCAATGATGATCA 784
|||||
seq_name: gb-est2:BI101730
seq_documentation_block:
LOCUS BI101730 1007 bp mRNA EST 26-JUN-2001
DEFINITION 602887560F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5042751
VERSION 5, mRNA sequence.
ACCESSION BI101730.1 GI:14552623
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 1007)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11117 row: h column: 16
High quality sequence stop: 840.
Location/Qualifiers
1..1007
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5042751"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 286 a 251 c 161 g 309 t
ORIGIN

```

```

alignment_scores:
Quality: 861.50 Length: 306
Ratio: 3.405 Gaps: 5
Percent Similarity: 82.680 Percent Identity: 57.190
alignment_block:
US-09-712-768-2 x BI101730
Align seg 1/1 to: BI101730 from: 1 to: 1007
193 ITeMetGlyAlaIleAsnMetIleThrThrPheLeuAsnMetArgAlaP 209
|||||
1 ATTTAGGTGCAATTAATTATTTATACCATATATATCAACATGAAGACCC 50
209 oGlyMetThrLeuHisLysValProLeuPheSerThrPheIlePheIle 226
|||||
51 AGCCATACACAGTATCAACTCCACTATTGTCGATCCGACTTATTA 100
226 hrAlaThrLeuIleLeuLeuAlaLeuProValLeuAlaGlyAlaIleThr 242
|||||
101 CAGCGGTACTGCTCCTATATACACAGAGTCTAGCGGAGGACTTACT 150
243 MetLeuLeuThrAspArgAsnPheGlyThrThrPhePheAsnProAlaG 259
|||||
151 ATACTACTAACAGACCGCACTTAACACAACTTCTTGTGATCCCGCTGG 200
259 yGlyGlyAspProIleLeuTyrGlnHisIleLeuThrPhePheGlyHisP 276
|||||
201 AGGAGGGACCAATCTCTTACCAGCATCTGTTGATCTTTGGGACC 250
276 roGluValTyrIleIleIleLeuProGlyPheGlyIleIleSerHisVal 292
|||||
251 CAGAGTTATATCTTATATCCTCCAGAGTTTGAATATATTACATGTA 300
293 ValSerThrPheSer...LysLysProValPheGlyTyrLeuProMetVal 308
|||||
301 GTTACTTACTACTCGGAAGAAAAGAACCTTGGGCTTATATGAAATGT 350
308 IYrAlaMetValAlaIleGlyValLeuGlyPheValAlaTyrAlaHis 325
|||||
351 ATGAGAAATATATGTCATTTGGCTTCTAGGCTTATATGATAGAGCCACC 400
325 IAsMetTyrThrValGlyMetSerLeuThrGlnGlnSerTyrPheMetLeu 341
|||||
401 ACATATTCACAGTAGAGATTAGATGATGACACAGAGCTTACTTACATCA 450
342 AlaThrMetValIleAlaValProThrGlyIleLysIlePheSerThrP 358
|||||
451 GCCACTATATATTCGCAATTTCTTACCGGTGCAAGATATTTAGCTGACT 500
358 eAlaThrMetTyrGlyGly...SerValGluPheLysSerProMetLeuThr 374
|||||
501 TGCACACCTTACACGAGGATTAATATGATCTCCAGCTATACTAGA 550
375 AlaPheGlyPheMetSerPheLeuPheThrValGlyGlyValThrGlyIle 391
|||||
551 GCCTTAGGCTTATTTCTTATTACAGTGGTGGCTTAACCGGAATGCT 600
391 ILeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyrTyrVal 407
|||||
601 TTATTCACACTATCCCTTGACATGCTGCTTACCATATCATATCTATGTA 650
408 ValAlaIlePheHisTyrValMetSerLeuGlyAlaIlePheAlaIleP 424
|||||
651 GTAAGCCATTTCCACATATGTTCTATCAATAGGAGCAGCTTTCATATCAT 700
424 eAlaGlyIleTyrPheTyrMetProLys...PheSerGlyAlaGlyAlaPhe 440
|||||
701 AGGAGATTTGTTCACTGATCCCATATATTTTCAGGCTTACCCCTAGA 750
440 oGluTyrAlaIleAlaLysLeuHisPheThrThrPhePheIleGlyAlaAsn 457
|||||
751 TGACCATTTGAGCAAAAGCCACTTGCATCATATTTCTGATAGGAAGTAACA 800

```

```

457 a1rhrphehepProgin.HisphleuGlyArgInGlyMetProlArgar 473
      :::::::::::::::::::: ||| ::::::::::::::::::::
801 TAACATCTTCCTCAACATCTTCCTGGGCTTACAGGATACCCGAGC 850
473 grryllleasprProGlnAlaPhlealeuTrrpAnLysValSerSert 490
      ||| :::::::::::::::::::: |||
851 CTACAGATACAGATGCTTACACACATTAACACTGCTCTCCTA 900
490 yrcGlyAlaPhe 493
      ||| |||
901 TAGGACATTTT 911

seq_name: gb_est1:AUI69837

seq_documentation_block:
LOCUS AUI69837 809 bp mRNA EST 29-JAN-2001
DEFINITION AUI69837 OL-br-ad cDNA Oryzias latipes cDNA clone br5288, mRNA
sequence.
ACCESSION AUI69837
VERSION AUI69837.1 GI:12591906
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE Mita,K., Ishikawa,Y. and Yamauchi,M.
1 (bases 1 to 809)
Establishment of cDNA database of medaka, Oryzias latipes
unpublished (2001)
AUTHORS Contact: Mita K
JOURNAL Genome Research Group
COMMENT National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsue@irs.go.jp
method:uni-directional sequence direction:sequenced from T3 primer
(5' -> 3').

FEATURES
source 1..809
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="br5288"
/clone_lib="OL-br-ad cDNA"
/sex="Female/male mixed"
/tissue_type="brain"
/dev_stage="adult"

BASE COUNT 203 a 193 c 153 g 259 t 1 others
ORIGIN

alignment_scores:
Quality: 858.50 Length: 265
Ratio: 3.700 Gaps: 1
Percent Similarity: 87.547 Percent Identity: 61.887

alignment_block:
US-09-712-768-2 x AUI69837 ..

Align seg 1/1 to: AUI69837 from: 1 to: 809

222 SerIlePheIleThrAlaTrpLeuIleLeuAlaLeuProValLeuAl 238
      :::::::::::::::::::: ||| ::::::::::::::::::::
11 GCCGCTACTATCACCGCAGATATTACTTCTTCTCTCTCCCTGTGCTAGC 60
238 aglAlaIleIleThrMetLeuLeuThrAspArgAnPheGlyThrThrPheP 255
      ||| :::::::::::::::::::: |||
61 TGCAGGATACATATCTTCTTACAGACCGAACAACCTAAATACACCTTCT 110
255 heasnrProAlaGlyGlyAspProIleLeuTrrpGlnHisIleLeuTrrp 271
      ||| :::::::::::::::::::: |||

```

```

111 TGCACCTCGACGAGAGAGGAGGCCATATCTCTACACACCTTTCGTA 160
272 PhePheGlyHisSprGluValTrrIleIleLeuProGlyPheGlyI 288
      :::::::::::::::::::: ||| ::::::::::::::::::::
161 TTTCTTGGCACCCCTGAAGTTTATTTCTATTTCTCCCGGCTGGAGAT 210
288 eileSerHisValValSerThrPheSer...LysLysProValPheGlyT 304
      ||| :::::::::::::::::::: |||
211 GATTTCATATTTGTGGCTTATTTATTTACAGTAAGAAAAGAACCTTGGCT 260
304 yIleuPromeValTrrAlaMetValAlaIleGlyValLeuGlyPheVal 320
      ||| :::::::::::::::::::: |||
261 ACATGGGAATGTATGAGCATGATGCTATCGGACTTCTAGGCTTATT 310
321 ValTrrAlaHisMetTrrThrValGlyMetSerLeuThrGlnGlnse 337
      ||| :::::::::::::::::::: |||
311 GTTGGACTCACCATATTTACCGGTAGGATGACGTGGACACTCGAGC 360
337 rTrrPheMetLeuAlaThrMetValIleAlaValProThrGlyIleLysI 354
      ||| :::::::::::::::::::: |||
361 TTATTTTACATCTCGACACATATTCATTCGATTCCTAGCTGGGTTAAG 410
354 lePheSerTrpIleAlaThrMetTrpGlyLysValGluPheLysSer 370
      ||| :::::::::::::::::::: |||
411 TATTTAGCTGACTCTGCTACCTGCTGATGAGTTCAATTAATGAGAAAC 460
371 PromeLeuThrAlaPheGlyPheMetPheLeuPheThrValGlyGly 387
      ||| :::::::::::::::::::: |||
461 CCACATATTATGGGCGCTGGTTATTTCTCTTTCACAGTGGGGGATCT 510
387 lrrhGlyIleValLeuAlaGlnAlaGlyLeuAspArgAlaTrrHisSpr 404
      ||| :::::::::::::::::::: |||
511 AACTGGCATGTTGCTTAGCCAACTCATCTTAGATATATCTGATGATACA 560
404 hTrrTrrValValAlaHisPheHisTrrValMetSerLeuGlyAlaIle 420
      ||| :::::::::::::::::::: |||
561 CTACTATTTAGTAGAGCTCTCCACTATGCTCTATCAATATAGGGCGGTA 610
421 PheAlaIlePheAlaGlyIleTrrPheTrrMetPrrLysPheSerGly 437
      ||| :::::::::::::::::::: |||
611 TTGCGCTTATATGGGGCATCTGTTCACTGATCCCTTATTTTCGGGCTA 660
437 gaIaPheProGluTrrAlaAlaLysLeuHisPheTrrThrPhePheIleG 454
      :::::::::::::::::::: |||
661 TACTTTACAGGAAACATGACAAAGATCCACTTGGAGTAAATTTGTGC 710
454 lylAlaSnValIleThrPhePheProGlnHisPheLeuGlyArgInGlyMet 470
      ||| :::::::::::::::::::: |||
711 GGGTAATTTTAACCTTCTT.CCTCAACACTTCTTAGGCTTAGCGGAGATA 759
471 ProArgArgTrrIleAspTrrProGlnAlaPheAlaLeuTrrpAn 485
      ||| :::::::::::::::::::: |||
760 CCACGACGATACTCTGACTACCCAGACGCTTATACCTATGAAAC 804

seq_name: gb_est1:BE468572

seq_documentation_block:
LOCUS BE468572 911 bp mRNA EST 28-JUL-2000
DEFINITION Iphak00346 Head kidney cDNA library Ictalurus punctatus cDNA
similar to cytochrome c oxidase subunit I, mRNA sequence.
ACCESSION BE468572
VERSION BE468572.1 GI:9559063
KEYWORDS EST.
SOURCE channel catfish.
ORGANISM Ictalurus punctatus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Siluriformes; Ictaluridae; Ictalurus.
REFERENCE Cao,D., Kocbas,A., Ju,Z., Karsi,A., Li,P., Patterson,A. and Liu
Z.J.
Transcriptome analysis of channel catfish (Ictalurus punctatus):
Gene cataloging and profiling from the head kidney

```





Thu Dec 27 08:21:37 2001

us-09-712-768-2.rst

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 10:54:55 ; Search time 39.94 Seconds  
(without alignments)  
1033.020 Million cell updates/sec

Title: US-09-712-768-2  
Perfect score: 3000  
Sequence: 1 MADAAIHGHHEKQGFETR.....PAHTFETLPKRDWDKHPSH 557

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3000	100.0	557	22	AAV97750
2	1169.5	39.0	581	22	AAAG92515
3	1160	38.7	552	22	AAAB79569
4	1078.5	36.0	648	22	AAAG81983
5	1010.5	33.7	663	22	AAAG98862
6	552.5	18.4	283	22	AAAB79571
7	340.5	11.3	204	22	AAAB79572
8	333.5	11.1	198	21	AAAB41481
9	303	10.1	56	22	AAV97762
10	186	6.2	472	21	AAAB26913
11	171	5.7	375	18	AAW20731

12	145.5	4.8	38	21	AAAB7168	Human prostate can
13	145.5	4.8	38	21	AAAB44050	Human cancer assoc
14	136	4.5	95	22	AAAB9570	Corynebacterium g1
15	136	4.5	377	22	AAU03645	Group B Streptococ
16	130.5	4.3	543	22	AAAG91478	C glutamicum prote
17	125.5	4.2	660	21	AAAT74602	Neisseria meningit
18	123	4.1	1019	22	AAAG92745	C glutamicum prote
19	121.5	4.0	418	22	AAAG91830	C glutamicum prote
20	120.5	4.0	660	21	AAAT74601	Neisseria meningit
21	116.5	3.9	1962	20	AAAT17250	NaMg polypeptide
22	115.5	3.9	386	22	AAAB79926	Corynebacterium g1
23	115	3.8	492	12	AAAT1360	Glucose transport
24	113	3.8	559	21	AAAB15974	E. coli proliferat
25	112.5	3.8	436	22	AAAG90123	C glutamicum prote
26	111	3.7	337	22	AAAG92904	C glutamicum prote
27	109.5	3.6	595	22	AAAG93189	C glutamicum prote
28	109	3.6	492	18	AAAT7835	Human glucose tran
29	108	3.6	452	22	AAAG93214	C glutamicum prote
30	108	3.6	452	22	AAAB76708	Corynebacterium g1
31	108	3.6	492	17	AAAB93000	Human G1U1 protei
32	107.5	3.6	484	22	AAAB76735	Corynebacterium g1
33	107	3.6	604	21	AAAT75522	Neisseria meningit
34	106.5	3.5	384	20	AAAT38852	Neisseria meningit
35	106.5	3.5	512	20	AAAT38556	Neisseria meningit
36	106.5	3.5	660	21	AAAT74600	Neisseria gonorrh
37	106.5	3.5	722	21	AAAT6163	Arabidopsis thalia
38	106	3.5	442	19	AAAB98870	H. pylori GHPo 171
39	106	3.5	442	19	AAAT71557	Helicobacter poly
40	106	3.5	442	22	AAAB6375	H. pylori HPN091 p
41	106	3.5	534	20	AAAT7288	Glucose transport
42	106	3.5	617	22	AAAB84532	Amino acid sequenc
43	106	3.5	617	22	AAAB84534	Amino acid sequenc
44	106	3.5	766	21	AAAT6161	Arabidopsis thalia
45	105	3.5	431	22	AAAG82736	S. epidermidis ope

## ALIGNMENTS

RESULT 1	
AAV97750	AAV97750 standard; Protein; 557 AA.
XX	XX
AC	AAV97750;
XX	XX
DT	06-NOV-2001 (first entry)
XX	XX
DE	G. oxydans cytochrome C oxidase (COI) protein sequence.
KW	Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;
KW	oxidative fermentation; electron transfer; respiratory chain; L-sorbose;
KW	2-keto-L-gluconic acid; 2KGA production; aldehyde production;
KW	carboxylic acid production; ketone production.
OS	Glucobacter oxydans.
XX	XX
PM	EPI103603-A2.
XX	XX
PD	30-MAY-2001.
XX	XX
PF	14-NOV-2000; 2000EP-0124785.
XX	XX
PR	17-NOV-1999; 99EP-0122842.
XX	XX
PA	(HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX	XX
PI	Asakura A, Hoshino T, Shinjoh M;
XX	XX
DR	WPI; 2001-357953/38.
XX	XX
DR	N-PSDB; AAA91489.
XX	XX
PT	New cytochrome c oxidase complex having cytochrome c oxidase activity
PT	from Glucobacter oxydans DSM 4025, useful in mediating electron

PT transfer in respiratory chain or producing 2-keto-L-gulononic acid from  
 L-sorbose or D-sorbitol  
 PS Claim 5; Page 16-20; 42pp; English.

CC This sequence is the Glucanobacter oxydans cytochrome C oxidase  
 CC (COI) of the invention. The COI complex is useful in improving oxidative  
 CC fermentation and is an essential component mediating electron transfer  
 CC in the respiratory chain. The recombinant microorganism and the  
 CC cytochrome C oxidase may be used in the genetic preparation of a  
 CC recombinant COI complex and in the production of 2-keto-L-gluconic acid  
 CC (2KGA) from L-sorbose or D-sorbitol in a culture medium. The COI is also  
 CC useful as a terminal oxidase, oxidizing cytochrome C, an electron  
 CC acceptor from an enzyme belonging to dehydrogenase for the production of  
 CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,  
 CC especially the production of 2KGA from L-sorbose or D-sorbitol.

XX Sequence 557 AA;

Query Match 100.0%; Score 3000; DB 22; Length 557;

Best Local Similarity 100.0%; Pred. No. 2,7e-293; Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MADAAIGHDHEKOGFTFRFMSSTNHKDIGLLYLVAAGVGFISLFTVMRLMDPG 60

1 madaaighdhekgoftrfmsstnhkdigllylvagvgfislftvmrlemdpg 60

61 VOYWCLEGARLIDASQTCANGHLWVTVYHGILMMFPVGPALFGGNGNLTMLDGT 120

61 vqywclegarlidasqtcanghlwvtyvtyhgilmfpvgpalfggngnltmldgt 120

121 APDMAFRMNNLSFWLFIAGTAMGVASLEAPGSGDGLSGSGVWVLYPPISTREAGYSMDL 180

121 apdmafrmnlnsfwlfiagtamgvasl eapggdglsgsgvswvlyppistreaagsmdl 180

181 AIPAVNHSGASSIMGATNMTTFLNMRARPGMTLHKVPLFMSFTITANLILALPLAAGA 240

181 aipavnhsgassimgatnmttflnmrarpgmtl hkvplfmsftitanlilalplaga 240

241 ITMLTDRNGCTFFPNFAGGDBILYOHILMPFGHPEVYIIIPGFIISHVSTSKRP 300

241 itmltdrngctffpnfaggdbilyohilmpfghpevyiilpgfiishvstskrp 300

241 itmltdrngctffpnfaggdbilyohilmpfghpevyiilpgfiishvstskrp 300

301 VEGYLPWVYMAVAIGVLGEVYVMAHMYTVGMSLTQSYEMLAIVTAIVPTGIKFSWIAF 360

301 veylpmvymavai gvlgevyvmaahmytv gmsltqsyemlaivtaivptgikfswiaf 360

361 MMGGSVEFKSPMLMARGFMLFTVGVGTGIVLAOAGIDRAVYHDTYVVAHFHYVMSLAI 420

361 mmggsvefkspmlmargfmlftvgvgtgiva oagidravyhdtvyvva hfyvmslai 420

421 FAFAGIYFMPKESGRAPPEMAAKLHFWTFICANVTFEPQHLRGQCPKRYIDYPPA 480

421 fafagiympek esgrappegmaaklhfwtfi canvtfepqhlrgqcpkryidyppa 480

481 PALMNKVSIGARLAPASFLFTVIVVYTLVAGRETRPNPWSGFADTLEWTLPSPPAH 540

481 palmnkvsigarlapasflftvivvytlvagr etrpnpwsgefadtlewtlpsppah 540

541 TFEETLPKRSMDKHPSH 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

541 tfeetlpkrsmdkhpsh 557

DE C glutamicum protein fragment SEQ ID NO: 6269.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

XX EPI108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOWA ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX N-PSDB: AAH67734.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 6269; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX Sequence 581 AA;

Query Match 39.0%; Score 1169.5; DB 22; Length 581;

Best Local Similarity 39.7%; Pred. No. 4.9e-109;

Matches 221; Conservative 110; Mismatches 188; Indels 37; Gaps 11;

12 HKKOGFTFRWEP--STNHKDIGLLYLVAAGVGFISLFTVMRLMDPGVQWMCLEGA 69

12 hkkogftfrwep--stnhkdigllylvagvgfislftvmrlemdpgvqwmclega 69

19 harkg-skawlmtdtkdqiglmymisfiffiggimalliaaelftpglqf----- 72

70 RLIDASQTCANGHLWVTVYHGILMMFPVGPALFGGNGNLTMLDGT 119

70 rli dasqtcanghlwvtyvtyhgilmfpvgpalfggngnltmldgt 119

130 NNLSFWLFIAGTAMGVASLEAPGSGDGLSGSGVWVLYPPISTREAGYSMDLAFVHL 187

130 nnlsfwlfiagtamgvasl eapggdglsgsgvswvlyppistreaagsmdlafvhl 187

120 naifgwltvvgvamlgflpgg----aafgwmvmsplsdahspglsgdwmlyvgva 175

168 SCASSIMGATNMTTFLNMRARPGMTLHKVPLFMSFTITANLILALPLAAGATMLTD 247

168 scassimgatnmttflnmrarpgmtl hkvplfmsftitanlilalplagaatmltd 247

176 tqigsvasaimlftlclirapgmctmtpiftnlftvsvlailfp11laaalyld 235

248 RNFQTFPNPAGGDPILYOHILMPFGHPEVYIIIPGFIISHVSTSKRPVGEVLP 307

248 rnfqtfpnpaggdpilyohilmpfghpevyiilpgfiishvstskrpvgevlp 307

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

236 rklgghlydpanggs-llwqhlftfignpevyialpifglivselipvriskpmtgyvgi 294

QY	308	YVAVNAVLIGLVGVYAAHMHMYTAMSLTQGSYFMLADPVIYAVPTGKIFSWLATIMGGSVE	367
Db	295	lfaclstlgaismawahmlnvgavll--pfIsfmcflIsVpCykfltnwvglmkxhlc	352
QY	368	FKSPMLAIFGFMLEFLTVGCVTGLVLAQAQGLDRAYHDTQYVVAHNFHYMSLGAIPFAIGT	427
Db	413	yIwFpRkntgmmderlqkhlfwlftlvgfngtlIshwgmngmprryadyIdsgfItyn	472
QY	486	KVSSYGAFILFASELEFIVFYVTVLVAQRERPRMPGSEFADILEWTLPSPPAHNPEETL	545
Db	473	qIstvfseflIglswlrlpImknvfkswrygelvtvddpwg-ygnslwatscprphnfaSl	531
QY	546	PK---RSDMDKHPSH	557
Db	532	pItrserpatelhyph	547
RESULT 3			
ID	AA879569	standard; Protein; 552 AA.	
AC	AA879569;		
XX			
DT	30-APR-2001	(first entry)	
DE		Corynebacterium glutamicum SMP protein sequence SEQ ID NO:654.	
XX			
XX		Corynebacterium glutamicum; carbon metabolism and energy production;	
XX		SMP protein; sugar metabolism and oxidative phosphorylation protein;	
XX		fine chemical production; organic acid; prokaryotic amino acid;	
XX		nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;	
XX		nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;	
XX		cardiolipin; aromatic compound; vitamin; cofactor; polypeptide; enzyme	
XX		diagnosis; Corynebacterium diptheriae; evolutionary study.	
OS		Corynebacterium glutamicum.	
XX			
XX		MO200100844-A2.	
PN			
XX			
PD			
XX	04-JAN-2001.		
XX			
XX	23-JUN-2000; 2000MO-IB00943.		
XX			
PR	25-JUN-1999; 9905-0141031.		
PR	08-JUL-1999; 99DE-1031412.		
PR	08-JUL-1999; 99DE-1031413.		
PR	08-JUL-1999; 99DE-1031419.		
PR	08-JUL-1999; 99DE-1031420.		
PR	08-JUL-1999; 99DE-1031424.		
PR	08-JUL-1999; 99DE-1031428.		
PR	08-JUL-1999; 99DE-1031431.		
PR	08-JUL-1999; 99DE-1031433.		
PR	08-JUL-1999; 99DE-1031434.		
PR	08-JUL-1999; 99DE-1031510.		
PR	08-JUL-1999; 99DE-1031512.		
PR	08-JUL-1999; 99DE-1031634.		
PR	08-JUL-1999; 99DE-1031840.		
PR	08-JUL-1999; 99DE-1032227.		
PR	08-JUL-1999; 99DE-1032230.		
PR	08-JUL-1999; 9905-0143208.		
PR	14-JUL-1999; 99DE-1032524.		
PR	14-JUL-1999; 99DE-1032973.		
PR	27-AUG-1999; 99DE-1040765.		
PR	31-AUG-1999; 9905-0151572.		
PR	03-SEP-1999; 99DE-1042076.		
PR	03-SEP-1999; 99DE-1042079.		
PR	03-SEP-1999; 99DE-1042086.		
PR	03-SEP-1999; 99DE-1042087.		

```

PR      03-SEP-1999;    99DE-1042088.
PR      03-SEP-1999;    99DE-1042095.
PR      03-SEP-1999;    99DE-1042123.
PR      03-SEP-1999;    99DE-1042125.
XX      03-SEP-1999;    99DE-1042125.
PA      (BAD1 ) BASF AG.
XX
PI      Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauser G;
DR      WPI: 2001-061975/07.
XX      N-PSDB: AAF71686.
PT      New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT      metabolism and oxidative phosphorylation protein for production or
PT      modulation of production of fine chemicals e.g. amino acids,
PT      carbohydriates or enzymes -
XX
PS      Claim 20; Page 1072-1074; 1246pp; English.
XX
AAAF1360 to AAFF71750 encode the Corynebacterium glutamicum sugar
CC      metabolism and oxidative phosphorylation (SMP) proteins given in
CC      AAB929243 to AAB 79633 which are involved in carbon metabolism and
CC      energy production. The C. glutamicum SMP gene can be used in vectors
CC      (II) for expression in host cells and production or modulation of
CC      production of fine chemicals, such as, an organic acid, a proteinoenic
CC      or nonproteinoenic amino acid (preferred), a purine or pyrimidine base,
CC      a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC      acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC      cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC      (III) encoded by them are used for diagnosing the presence or activity of
CC      Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
CC      containing them are used to map genomes of organisms related to
CC      C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC      in evolutionary studies, in determining SMP protein regions required
CC      for function, in modulating SMP protein activity, in modulating the
CC      metabolism of sugars, and in modulating high-energy molecule production
CC      in a cell (I.e. ATP, NADPH).
XX
SQ      Sequence      552 AA:

Query Match          38.7%: Score 1160; DR 22; Length 552:
Best Local Similarity 40.0%; Pred No. 4.2e-108:
Matches 217; Conservative 107; Mismatches 184; Indels 34; Gaps 9.

   24 SPNKKDGLLTVAAAGVGSEISVLTFTVMRLLEMDRGVOYMCLEGARLLADASQTCFANG 83
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      3 tcdhkqglmylmstfssffgllmalrraeifrgyql-----sn 45

   84 HLMMNVVTYHNGILMFMFVFGIRALFGSGFYVLMPLQLGADMDAFAVRNNNSPMLFLAGTGM 143
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      46 eqnglfmbmgtvmllylrbtw-gfanyvlrldgadrvfrplnafgtltvgava 104

   144 GVASLPARGGDGLSGGVWLVLRPLS--TREAGYSMDLAIFAHNLSGASSINGAINMIT 201
       |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      105 mlgtfltrpg---aadfwlmsysprlsdeihbpdlgsdmwlygvatgytsvaasainmlt 160

   202 TFLNMRAFGATLKRVLFESMSFTTAMLLALPVLAGAITMLLDNRNGTTFPNPAGG 261
       |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      161 flclrtarpmtemfmplrfwnlifvavslallfprllaaagvlydrkignrlhydpangg 220

   262 DPLTVHHILPFCHRNRYLIIPGFETIHVVSTFSKRPVFCGLPMVAVMAVIGVCVFV 321
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      221 s-tlwqhlltwfgrhevuyalrpfisigiveellprvsrkpmfyggvllftalsigaismay 279

   322 WAHMNTVGSMSLTQOOSYFMLATMVIAVPTGIKTSNIATIMWGGSVEFKSPMLWAEGFMEL 381
       |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      280 walmntvtgavll-pffsfmctflisvrpykvkfhwgmvmxghiltwecmtlwsyvfmat 337

   382 FYWGGVTGTIVLAQADRAKHDTYYVAAPHNYMSLGALFALFAGTYTFMPREFSGAFAPE 441
       |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      338 flfgytltgtmlaspriidfnadasyfliathbytlfgtvlvafascagyfwrpkmtygmmdc 397

```

**Qy** 442 MAAKLHEFFEEFGAVNTEPFOEFLGRGOWPRYIDY--PEARALMNKYSVGAFLAFASEF 499  
|::||: |: : || |:::| ||||| |: : |: ::::|: |:  
**Dd** 398 rlgkhhfwlftvgfngftllgwnvgnmgpriradyldsdgtlyndqstlvfyllglsv 457  
: |: : |: : |: : ||| |: : |||: |: :  
**Oy** 500 LEFIVIEVTYLVAQRRETRPNMGEEADLTLEWTLPSPPAHTEFTLPK----RSDMDKNP 555  
: |: : |: : |: : ||| |: : ||| |: : |||: |: :  
**Dd** 458 lpfivwvfksrwtygelvtvdwdpw-ygnslawatcsprrhnfaelplrseerpfelh 516  
556 SH 557  
|  
**Dd** 517 ph 518

**RESULT** 4  
**AAG81983**  
**ID** AAG81983 standard; Protein; 648 AA.  
**XX** AAG81983;  
**AC**  
**DF** 03-SEP-2001 (first entry)  
**DE** S. epidermidis open reading frame protein sequence SEQ ID NO:1060.  
**XX**  
**KM** Staphylococcus epidermidis SRI strain; infection; diagnosis;  
**KX** vaccination; endocarditis.  
**OS** Staphylococcus epidermidis.  
**PN** WO200134809-A2.  
**XX** 17-MAY-2001.  
**PD**  
**PF** 09-NOV-2000; 2000WO-US30782.  
**XX**  
**PR** 09-NOV-1999; 99US-0164258.  
**XX**  
**PA** (GLAX ) GLAXO GROUP LTD.  
**XX**  
**PI** Kimerly WJ;  
**DR** WP1: 2001-316495/33.  
**N-PSDB:** AAH52833.  
**PT** Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
**XX** useful for vaccinating against infections, e.g. endocarditis -  
**Claim 18;** Page 309; 2188pp; English.

**AAH52304** to **AAH53970** represent nucleic acids (I) encoding polypeptides (II), given in **AAG81454** to **AAG83120**, from *Staphylococcus epidermidis*, (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. **AAH53971** to **AAH55090** represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. **AAH55091** to **AAH55098** represent oligonucleotide sequences and primers which are used in the exemplification of the present invention.

**N.B.** The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

**Sequence** 648 AA;

**Query Match** 36.0%: Score 1078.5; DB 22; Length 648;

[illegible]



PS Claim 20; Page 1076-1077; 1246pp; English.

XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins  
 CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein sequences required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).

XX  
 SQ Sequence 283 AA:

Query Match 18.4%; Score 552.5; DB 22; Length 283;  
 Best Local Similarity 41.3%; Pred. No. 2.4e-47;  
 Matches 104; Conservative 47; Mismatches 92; Indels 9; Gaps 4;

OY 312 VAIGVLGFFVVAHMHVTVGMSLQOQSYFMLATMVIAVPGIKIFESMTATMGSSVEKSP 371  
 Db 1 ISIGALSMawahmTVtgvll--pLsfmctllsvrGvKtltvGvmKvKhltwep 58  
 OY 372 MLNAFGMFLFTVGTGLVLAQAGLDRAVHDYVVAHNVMSGLAFALFAGTIFYM 431  
 Db 59 Mlsvsgfmatflfgltglnlaasprldfnladsyfliahbhtlftvtvfaasgtyvlf 118  
 OY 432 PKRSGRAFPRMAKLNHFVTFITGANVTFPQHLGRQGMRRITDY--PEAFELMKVSS 489  
 Db 119 pKntgmmderlkgkhtlftvgfhgtflghvgnmgnprtyadyldsdgftlyngist 178  
 OY 490 YGAFLEAFSLFFVTVFVVTVAGRRETRPNMGCEFADTLEWLPSPPAHPEETLPK-- 547  
 Db 179 vYflllglyslpflmvfkswtygelvtvdpwg-ygnslwatsecpplmhasiprlr 237  
 OY 548 --RSDMDKHPSH 557  
 Db 238 serpafelhyph 249

RESULT 7

AAB79572 ID AAB79572 standard; Protein; 204 AA.

XX AAB79572;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:660.

XX Corynebacterium glutamicum; carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX Corynebacterium glutamicum.

XX WO200100844-A2.

XX 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB00943.

XX  
 PR 25-JUN-1999; 990S-0141031.  
 PR 08-JUL-1999; 99DE-1031412.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031431.  
 PR 08-JUL-1999; 99DE-1031433.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031562.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 09-JUL-1999; 99DE-1032180.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 09-JUL-1999; 990S-0143208.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 31-AUG-1999; 990S-0151572.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042099.  
 PR 03-SEP-1999; 99DE-1042123.  
 PR 03-SEP-1999; 99DE-1042125.

XX (BADI ) BASF AG.

XX Pompejus M. Kroege B. Schroeder H. Zelder O. Haberbauer G;

XX WPI: 2001-061975/07.

XX N-PSDB: AAF71689.

XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes -

PS Claim 20; Page 1079; 1246pp; English.

CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins  
 CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).

XX Sequence 204 AA;

Query Match 11.3%; Score 340.5; DB 22; Length 204;  
 Best Local Similarity 41.6%; Pred. No. 3.4e-26;  
 Matches 67; Conservative 31; Mismatches 56; Indels 7; Gaps 3;





XX Example 3; Page 28; 42pp; English.

CC This sequence represents a fragment of the Gluconobacter oxydans  
CC cytochrome C oxidase, COI. The invention relates to a cytochrome C  
CC oxidase COI complex. The COI complex is useful in improving oxidative  
CC fermentation and is an essential component mediating electron transfer  
CC in the respiratory chain. The recombinant microorganism and the  
CC cytochrome C oxidase may be used in the genetic preparation of a  
CC recombinant COI complex and in the production of 2-Keto-L-gluconic acid  
CC (2KG) from L-sorbose or D-sorbitol in a culture medium. The COI is also  
CC useful as a terminal oxidase, oxidizing cytochrome C, an electron  
CC acceptor from an enzyme belonging to dehydrogenase C, for the production of  
CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,  
CC especially the production of 2KG from L-sorbose or D-sorbitol.

XX Sequence 56 AA:

Query Match 10.1%; Score 303; DB 22; Length 56;  
Best Local Similarity 98.2%; Pred. No. 3.8e-23;  
Matches 55; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 271 WFGHPEVYIIILPGFGLISHVSTFSKRPVGYLPYAMVAIGVLGFVVAHHM 326  
DB 1 wffghpevyiilpgfglisshvstfskkrpvfgylpmyamlaigvlgfvvahn 56

RESULT 10

AAB26913 AAB26913 standard; Protein; 472 AA.

AC AAB26913;

DT 12-JAN-2001 (first entry)

DE Large subunit of nitric oxide reductase NorB.

KM NorB; nitric oxide reductase; enzyme; transgenic organism;

KM toxic waste denitrification; denitrifying bacteria.

OS Pseudomonas sp. G-179.

PN US6136588-A.

PD 24-OCT-2000.

PF 15-JUL-1999; 99US-0354129.

PR 17-JUL-1998; 98US-0093191.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

PI Bedzyk LA, Ye RW;

DR WPI: 2000-686079/67.

DR N-PSDB; AAA94683.

PT New nucleic acid encoding bacterial nitric oxide reductase, useful e.g.  
PT for preparing transgenic organisms able to detoxify wastes

PS Claim 1; Columns 21-24; 24pp; English.

CC The present sequence is the NorB protein from Pseudomonas sp. strain  
CC G-179. This sequence is the large subunit of nitric oxide reductase. The  
CC norB gene can be used to create transgenic organisms that can denitrify  
CC toxic waste. In addition, the norB gene may be used to produce probes and  
CC primers for identification of new denitrifying bacteria.

SO Sequence 472 AA:

Query Match 6.2%; Score 186; DB 21; Length 472;

Best Local Similarity 21.7%; Pred. No. 3.6e-10;  
Matches 98; Conservative 77; Mismatches 173; Indels 104; Gaps 23;

OY 89 WTYIGITLMFEVFGICPALFGEGGNTLMPLOIGAPDMPERNUNLSFWFLAGTAMGVASL 148  
DB 51 mhtnrltwwlllg---fmgaelyllp-eeaeetlyspklatagfwlfliaaiaav-- 103  
OY 149 FAPGDDQLGSGVGVLPPLSTREAGYSMDLAIFAVHLSGASSIMGAINMTFELMRA 208  
DB 104 -----gymfkinegreffieqfllkiglyiv-----clmfllnvtmstlgyrk 146  
OY 209 PGMTLHKVPLEFS-NSIFITAMLLALPVLGATMTMLTDNFGTTFEPNAG-GGDEPLY 266  
DB 147 tvvt-nlllglwglalf-----flsfypnaldldkmw 181  
OY 267 OHT--LMFEGHPEVYIIILPGF-----GIISHVSTFSKRPVGYLPYAMVAI---- 314  
DB 182 yvvvhlwvegwelmasvialflmklngldrevvkw-----lyvlgglalfs 230  
OY 315 GVLGFVVAHHMYTIVGM-SLTOQSYFMLATWIAVPTGIRKIFSMIATMW--GGSVEFKSP 371  
DB 231 gllg---tghlywyigapgywqwglsfslvleapfflmvffltvmt-wkagrkhprraa 286  
OY 372 MLMAFG-FMFLFTVGVGTIVLAQAGLDRAVHDITYVVAHF-----YVNSLGATPAIF 424  
DB 287 llwstgcsvmafffgqvwgflhtlssvnychgvtahghlaffgaymnlalmaa-- 344  
OY 425 AGIYFYMRFSGR-AFPWMAKLIHFWTFEGANVTFFPQHLG-----RQGMRRRTIDY 477  
DB 345 ---yavpelcgkrpkynglsmvsfmmctamsvmltalftagylqvlgrvlggyndv 400  
OY 478 PEAFAL--MNKSSYGAFAPASFLFVIFV 507  
DB 401 qdqlalfywrzlgv-gvflvlsalmfwailv 431

RESULT 11

AAM20731 AAM20731 standard; protein; 375 AA.

AC AAM20731;

DT 16-JUL-1997 (first entry)

DE H. pylori inner membrane protein, 06cp1118orf6.

KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

OS Helicobacter pylori.

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; 96WO-US09122.

PR 01-APR-1996; 96US-0630405.

PR 07-JUN-1995; 95US-0487032.

PA (ASTR ) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI: 1997-052306/05.

DR N-PSDB; AAT67984.

PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) -- useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter

PS Claim 56; Page 1146-1147; 1481pp; English.

XX The present sequence is a Helicobacter pylori inner membrane protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.

XX Sequence 375 AA:

Query Match 5.7%; Score 171; DB 18; Length 375;  
 Best Local Similarity 22.2%; Pred. No. 8.7e-09;  
 Matches 75; Conservative 64; Mismatches 143; Indels 56; Gaps 17;

QY 222 SIFIAWLLALPYLAGAITMLTRNGTFFENPAG-----GCDPIIYQHILW 271  
 Db 48 Ilyslwyiat--yvglavmyifnlsipcyfvadngsvhsismysndalig--w 102  
 QY 272 FFGHEVYIIILPG-FGLISHVSTFSKRPVGYLPMYANVAIGVGFV-WMA--HHMY 327  
 Db 103 wghnaavfvtlsqvglyfilyfllpkessqplfsyKltflfws---lmfyilwagghhll 159  
 QY 328 TVGMSLTQGSYFMLATVIAVPTGIKIPSWIATMWSGSEVERK-SPMLMAFGFENL---F 382  
 Db 160 ystvdpwqqltssvsvslilpswgtalimlltmrgvqhkspll---kfvlastly 216  
 QY 333 TVGGVTGIVLAQAGIDRAHYDHYVVAHFHYVMSIGAI-FAIFAGIYFYMKR-----FS 435  
 Db 217 mlstlegisqiksvnaiahftdwilghvhdgv-lgwygflisamyhmprrlKkrelys 275  
 QY 436 GRAPEMAKALHEWTFEFGANVTFEPQHLG-ROGMRRP-----YIDPEAFAL 483  
 Db 276 gr-----lvdfqfwmrltgiylyfssmvlagltgmmwrdvdygnllygtidvkalip 330  
 QY 484 WNKVSSYGAFLEAFASFLFFIVFYVTLVAGRR-ETRPN 520  
 Db 331 yynlgyvgslmyfiglilfaynlmftltaqklerepn 368

RESULT 12

AAB57168  
 ID AAB57168 standard; Protein: 38 AA.

XX AAB57168;

XX 13-MAR-2001 (first entry)

DE Human prostate cancer antigen protein sequence SEQ ID NO:1746.

XX Human: prostate cancer; prostate cancer antigen; detection; diagnosis;  
 KW neuroprotective; cytoskeletal; cardioprotective; immunomodulatory; muscular;  
 KW vulnereary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 KW wound; infectious disease.

OS Homo sapiens.

PN WO200055174-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05988.

XX

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

DR WPI: 2000-587513/55.

DR N-PSDB; AAF16371.

XX Prostate cancer associated gene sequences, referred to as prostate  
 PT cancer antigens, useful for treatment, prevention, and diagnosis of  
 PT disorders such as prostate cancer -

XX Claim 11: Page 2216; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated  
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytoskeletal,  
 CC cardioprotective, immunomodulatory, muscular, vulnereary, gastrointestinal,  
 CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.

XX Sequence 38 AA;

Query Match 4.8%; Score 145.5; DB 21; Length 38;  
 Best Local Similarity 73.7%; Pred. No. 1.6e-07;  
 Matches 28; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 272 FFGHEVYIIILPGFGLISHVSTFS-KRPVFGYGLMV 308  
 Db 1 ffghevyiilpfglshvstfysgkkeep19yigmv 38

RESULT 13

AAB44050  
 ID AAB44050 standard; Protein: 38 AA.

XX AAB44050;

XX 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1495.

XX Human: cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytoskeletal; proliferative; vulnereary; immunomodulatory;  
 KW antidiabetic; antihistaminic; antineoplastic; antitubercular; antiviral;  
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; noctropic;  
 KW vasoregic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.

OS Homo sapiens.

PN WO200055350-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 10:57:21 ; Search time 21.29 Seconds  
(without alignments)  
588.743 Million cell updates/sec

Title: US-09-712-768-2

Perfect score: 3000

Sequence: 1 MADAAIHGDHHEKGFETR.....PAHTEPLPKRSDMDKHPSH 557

Scoring table: BLASTM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1531	51.0	513	4	US-09-097-889-15
2	186	6.2	472	4	US-09-334-129-2
3	108	3.6	492	2	US-08-355-84-3
4	108	3.6	492	2	PCT-US95-16126-3
5	106	3.5	534	2	US-09-031-392-4
6	106	3.5	534	4	US-09-299-549-4
7	106	3.5	617	1	US-07-879-617A-11
8	106	3.5	617	1	US-08-301-722A-3
9	106	3.5	617	1	US-08-240-783B-3
10	106	3.5	617	1	US-08-753-985-11
11	106	3.5	617	3	US-09-084-813-3
12	106	3.5	617	5	PCT-US92-09662-3
13	103	3.4	451	1	US-08-191-337-3
14	98.5	3.3	383	2	US-09-031-392-3
15	99.5	3.3	383	4	US-09-299-549-3
16	98.5	3.3	384	2	US-08-928-692-13
17	98.5	3.3	1956	4	US-08-843-417-10
18	97	3.2	506	4	US-09-066-047-7
19	97	3.2	662	1	US-07-841-651-4
20	96.5	3.2	524	2	US-08-928-692-12
21	96	3.2	493	2	US-09-031-392-10
22	96	3.2	493	4	US-09-299-549-10
23	95.5	3.2	708	2	US-08-576-165-2
24	95	3.2	326	4	US-08-986-768-2
25	94.5	3.1	503	4	US-09-068-195-24
26	93	3.1	422	4	US-09-625-188-12
27	92.5	3.1	460	2	US-08-677-049-10

28	91.5	3.0	369	1	US-07-937-609-19	Sequence 19, Appl
29	91.5	3.0	369	4	US-08-029-170-19	Sequence 19, Appl
30	91.5	3.0	384	2	US-08-103-170-10	Sequence 10, Appl
31	91.5	3.0	509	2	US-09-031-392-6	Sequence 6, Appl
32	91.5	3.0	509	4	US-09-299-549-6	Sequence 6, Appl
33	91	3.0	417	1	US-08-553-888A-3	Sequence 6, Appl
34	91	3.0	607	1	US-07-879-617A-12	Sequence 12, Appl
35	91	3.0	607	1	US-08-753-985-12	Sequence 12, Appl
36	91	3.0	685	4	US-08-960-048-7	Sequence 7, Appl
37	91	3.0	2265	2	US-08-149-097D-36	Sequence 36, Appl
38	91	3.0	2509	2	US-08-149-097D-35	Sequence 35, Appl
39	91	3.0	3169	2	US-08-477-451-6	Sequence 6, Appl
40	90.5	3.0	418	4	US-09-030-267-5	Sequence 5, Appl
41	90	3.0	397	1	US-08-098-141-2	Sequence 2, Appl
42	90	3.0	453	1	US-08-439-131A-5	Sequence 5, Appl
43	90	3.0	453	1	US-08-440-674-4	Sequence 4, Appl
44	90	3.0	635	2	US-09-014-969-11	Sequence 11, Appl
45	89.5	3.0	714	4	US-09-115-954-4	Sequence 4, Appl

## ALIGNMENTS

```

RESULT 1
US-09-097-889-15
: Sequence 15, Application US/09097889
: Patent No. 6218117
:
GENERAL INFORMATION:
: APPLICANT: Herrnschadt, Corrina
: APPLICANT: Ghosh, Soumitra S.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
: TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: City: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/097,889
: FILING DATE: 15-JUN-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Roseman Ph.D., Stephen J.
: REGISTRATION NUMBER: 43,058
: REFERENCE/DOCKET NUMBER: 660088.417
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 513 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
:
US-09-097-889-15

```

Query Match 51.0%; Score 1531; DB 4; Length 513;  
Best Local Similarity 52.4%; Pred. No. 5.2e-143;  
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

Db	2	FADRLFTFNKNDIGTLYLLFGAGAGVGLGATLSLLIRAEIGDRP-----	45
Qy	77	QCTANGNGLKAVVMTYHAGTILMFEEVGIAPALFGGCGNTLMPLDQAGPMAAPRRNNLSFWL	136
Db	46	-NLNDHIYVYIYTAHFAVIFMFVMEIMMGCGNMLVPLMIGAPMAPRRNNMSFWL	104
Qy	137	FLAGAMCVASLEAPGGDGOGLSGGVWLVLPSELR--EAGYSMDLAFVNHLSGASIM	194
Db	105	LPPSLILLASAMV-----EAGAGTGMTVYRPLAGNTHFGASVDLTTFSLHLAGVSSIL	159
Qy	195	GAIMNITTELMKRAPGMYLHKVPLFSSIFETAMLILALPVLALGATIMLTDNRNGETP	254
Db	160	GAINFTTITIMKPRAMTQYOTPLFVMSVLITAVILLLSLPLVALGITMLLTDNRNITFE	219
Qy	255	ENPAGGGRPLXONILMFEGHREYVYIILPFGGIIINHVSTFS-KKRVFGYLPVATAMVA	313
Db	220	FDPAGGGRPLXONILMFEGHREYVYIILPFGGIMSHIVYISGKKRPFYMGVWVAMMS	279
Qy	314	IGVLGVVAMNMYTVGSLTQDSYFMLATVNAVPPGIKFISLAIAMGSGSEFKSPML	373
Db	260	IGLGLFETVAMNMFVGDVDTYRPLFESATMTIALIPGVKVFSLATLHSSNMKMSAVL	359
Qy	374	WAFGMELETVGCVTGLVLAAGLDRAYHDTYVVAHFHYVMSLGAIFAIAGIYFYMPK	433
Db	340	WALGIFLEFVYGGTGLVLAASLDIYHDTYVVAHFHYVLSGAVAFALMGCFIMHPL	399
Qy	434	FSGRAFFPMAKNHEWTFETCANVTFEPPOHFLGQGRPRRTIDYPEAFALMKNVSSGAF	493
Db	400	FSGYLDDQYAKIHETIMFICVNLTFFPOHFLGSGMPRRYSIDYPTWVNTLLSSVGSF	459
Qy	494	LAFASLEFPIYIEVYTVLAVGRRETRPNMGFEAD-----TLEWTLPSPPAHT	541
Db	460	ISLTNVMLITMI-----WEAFASKRVYLVWEPSSNMLBMLYGCPPYHT	504
Qy	542	FE 543	
Db	505	FE 506	
RESULT	2		
US-09-354-129-2			
Sequence 2, Application US/09354129			
Patent No. 6136588			
GENERAL INFORMATION:			
APPLICANT: YE, RICK W.			
TITLE OF INVENTION: GENES ENCODING DENITRIFICATION REACTIONS			
FILE REFERENCE: CL-1255			
CURRENT APPLICATION NUMBER: US/09/354,129			
CURRENT FILING DATE: 1999-07-15			
EARLIER APPLICATION NUMBER: 60/093,181			
EARLIER FILING DATE: July 17, 1998			
NUMBER OF SEQ ID NOS: 14			
SOFTWARE: Microsoft Office 97			
SEQ ID NO 2			
LENGTH: 472			
TYPE: PRT			
ORGANISM: Pseudomonas sp. strain G-179			
US-09-354-129-2			

Query Match	6.28;	Score	186;	DB	4;	Length	472;
Best Local Similarity	21.78;	Pred.	No. 3.3e-10;				
Matches	98;	Conservative	77;	Mismatches	173;	Indels	104;
						Gaps	23

```

Oy      89  MVTYGIILMFEFVGIPIALFEGFGVYIMDLQIAGDMMFPRNNLSFWLTLFAGTAMVASTL 148
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      51  MHTNALIWLVLIG---FMGATYYLLP-EEAEHELKSPKALAQNFILFLAALIANV-- 103
Oy      149  FARGDQGLSGSGVGVNLYLPRLSTEEAGYSMDLAFVNHLSASSIMKAIMWITFLNMR 208
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      104  -----GYMKTHGEEFLEQPIILITGLIVY-----CLMFLFNWTMSLEGRK 146
Oy      209  PGMTLHKVPLEF-WSIFTLAWLILALPVLVAGATMLTLDRNQTTFENPAG-GGDPILY 266

```

```

Db      147 TVVI--NLIFGLMGIAIF-----LFSTSYNANLALDKMY 181
OY      267 OHI--LMFEGPEVYIIILPGF-----GLISHVSFFSKRVFGYLLPMYANVAI---- 314
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      182 WYVHLMEGWGVELIMASVLAFLMIKILNGIDREVEYK-----LYVIVGLAES 230
OY      315 GVLGVVVAHHMYVGM--SLQOSYVPLAIPVIAVPGIKIFSMIAIM--GGSVFEKSP 371
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      231 GILG---TGHYYIWIGADGYQWQMGSLPSTLEVAFFTWVITFEVMT--WAGKRIHPRAA 286
OY      372 MLMAFG--FMELFTYGGVYTGILQAQGLDRAYHDYVYVAHFH-----YVMSGALFAIF 424
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      287 LMSIGCSVMAFCGAGVWGFLHTLSSVNYTHGQVTAHGHILFCAGYVYMLNLAIMA-- 344
OY      425 AGIIFYMPKESGR--AFPEMAKLLHFWTFFGIANVFEPQIFLG-----RQMPKRIIDY 477
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      345 ----YAVPEIGRPRYQWMLSMVFMMKSTAMSVYFPAFLFAGLOYHLOIVLGEQMDV 400
OY      478 PEAFAL--WNKVSSYGAFIASFELFIVFV 507
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      401 QDQALFLYVWRIGS--GVDFVLSALMFYVAILV 431

```

```

1      RESULT 3
2      US-08-355-844-3
3      Sequence 3, Application US/08355844
4      Patent No. 5940307
5      GENERAL INFORMATION:
6      APPLICANT: Fischbarg, Jorge
7      APPLICANT: Czegledy, Ferenc
8      APPLICANT: Iserovich, Pavel
9      APPLICANT: Li, Jun
10     APPLICANT: Cheung, Min
11     TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
12     TITLE OF INVENTION: STRUCTURE
13     NUMBER OF SEQUENCES: 3
14     CORRESPONDENCE ADDRESS:
15     ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
16     STREET: 30 Rockefeller Plaza
17     CITY: New York
18     STATE: NY
19     COUNTRY: USA
20     ZIP: 10112-0228
21     COMPUTER READABLE FORM:
22     MEDIUM TYPE: Floppy disk
23     COMPUTER: IBM PC compatible
24     OPERATING SYSTEM: PC-DOS/MS-DOS
25     SOFTWARE: PatentIn Release #1.0, Version #1.25
26     CURRENT APPLICATION DATA:
27     APPLICATION NUMBER: US/08/355,844
28     FILING DATE: 14-DEC-1994
29     CLASSIFICATION: 436
30     ATTORNEY/AGENT INFORMATION:
31     NAME: Tang, Henry Y. S.
32     REGISTRATION NUMBER: 29,705
33     REFERENCE/DOCKET NUMBER: A29927-50/29910
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: 212-408-2586
36     TELEFAX: 212-765-2519
37     INFORMATION FOR SEQ ID NO: 3:
38     SEQUENCE CHARACTERISTICS:
39     LENGTH: 492 amino acids
40     TYPE: amino acid
41     STRANDEDNESS: single
42     TOPOLOGY: linear
43     MOLECULE TYPE: peptide
44     ORIGINAL SOURCE:
45     ORGANISM: Human
46     FEATURE:
47     NAME/KEY: Peptide
48     LOCATION: 1..492
49     OTHER INFORMATION: Facilitative glucose transporter

```



OTHER INFORMATION: Glut1 protein  
US-08-355-844-3

Query Match 3.6%; Score 108; DB 2; Length 492;  
Best Local Similarity 20.3%; Pred. No. 0.018;  
Matches 92; Conservative 57; Mismatches 160; Indels 144; Gaps 19;

QY LMNVV-----TYHGILMFEFGIPALFEGFGNYLMPLO-----IGAPDMAFPRMNN-----LS 133  
DB LMSLSVAIFSVGMIQSGSVGLFVNRFRGRNSMLMMLLAFVSAVLMGFSKLSFEMLI 123  
QY 134 FWLFIAGTAMGVASLFPAGDQGLSGVGVLYPLSTREA-----GYSMDLAFVHLS 188  
DB 124 LGRFTIGVCGITGTGFVPMVYGEVS-----PTAFRGALGTLHQLGIVGLIAQVF 174  
QY 189 GASSIMGAINMTTFLNMRAPGMTLHKVPLFSWISFTIAMLILLAPVLACATMLLDR 248  
DB 175 GLDSTMGKNDL-----WPLLSTIFIPALLOCIIVIPCPSPREFLLNR 218  
QY 249 N-----FGTT-----FNNAGGDPILVQHIL 270  
DB 219 NEENRAKSVLKRLGTADVTHDLOEMKEESRQMRKKVTIIEFRSPAYRQPIILAVVL 278  
QY 271 WFGHPREVYIIILPGFGIISHVSTFSKRPVGYLPMVYAMVAGVLG-----FVY- 321  
DB 279 QLSQO-----LSGINAVFYSTSIPEKA--GYOQPVYATIGSGIVNTAFVVSLEFVE 329  
QY 322 -----WAHMYVYVMSLQOQSYFMLATVIAVPTGIKIFSMIATMGSGVEKSPMLMAF 376  
DB 330 RAGRTLHLIGLAGMA-----GOAILMTIALAL--LEQLPMSYV-----SIVAI 373  
QY 377 GFMLFTVIG-----GVGTGIVLAQGLDRAVHDYVVAHFHYVMSLGA-- 419  
DB 374 GFVAFEFVGPGPIPMFTIVALESGQPRPAIIVAGFSN--WTSNFIYVGMCFQYVDELQGPY 432  
QY 420 IFALFAGI-----YFMPKFSGRAPEWAA 444  
DB 433 VFILFTVLVLFIRTYFKVPETKGRTFDEIAS 465

RESULT 4  
PCT-US95-16126-3  
Sequence 3, Application PC/TUS95J16126  
GENERAL INFORMATION:  
APPLICANT: Fischbarg, Jorge  
APPLICANT: Czegledy, Ferenc  
APPLICANT: Iserovich, Pavel  
APPLICANT: Li, Jun  
APPLICANT: Cheung, Min  
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10112-0228  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16126  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/355,844  
FILING DATE: 14-DEC-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Tang, Henry Y.S.  
REGISTRATION NUMBER: 29,705  
REFERENCE/DOCKET NUMBER: A29927-50/29910  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-408-2586  
TELEFAX: 212-765-2519  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Human  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..492  
OTHER INFORMATION: Facilitative glucose transporter  
OTHER INFORMATION: Glut1 protein  
PCT-US95-16126-3

Query Match 3.6%; Score 108; DB 5; Length 492;  
Best Local Similarity 20.3%; Pred. No. 0.018;  
Matches 92; Conservative 57; Mismatches 160; Indels 144; Gaps 19;

QY 85 LMNVV-----TYHGILMFEFGIPALFEGFGNYLMPLO-----IGAPDMAFPRMNN-----LS 133  
DB 64 LMSLSVAIFSVGMIQSGSVGLFVNRFRGRNSMLMMLLAFVSAVLMGFSKLSFEMLI 123  
QY 134 FWLFIAGTAMGVASLFPAGDQGLSGVGVLYPLSTREA-----GYSMDLAFVHLS 188  
DB 124 LGRFTIGVCGITGTGFVPMVYGEVS-----PTAFRGALGTLHQLGIVGLIAQVF 174  
QY 189 GASSIMGAINMTTFLNMRAPGMTLHKVPLFSWISFTIAMLILLAPVLACATMLLDR 248  
DB 175 GLDSTMGKNDL-----WPLLSTIFIPALLOCIIVIPCPSPREFLLNR 218  
QY 249 N-----FGTT-----FNNAGGDPILVQHIL 270  
DB 219 NEENRAKSVLKRLGTADVTHDLOEMKEESRQMRKKVTIIEFRSPAYRQPIILAVVL 278  
QY 271 WFGHPREVYIIILPGFGIISHVSTFSKRPVGYLPMVYAMVAGVLG-----FVY- 321  
DB 279 QLSQO-----LSGINAVFYSTSIPEKA--GYOQPVYATIGSGIVNTAFVVSLEFVE 329  
QY 322 -----WAHMYVYVMSLQOQSYFMLATVIAVPTGIKIFSMIATMGSGVEKSPMLMAF 376  
DB 330 RAGRTLHLIGLAGMA-----GOAILMTIALAL--LEQLPMSYV-----SIVAI 373  
QY 377 GFMLFTVIG-----GVGTGIVLAQGLDRAVHDYVVAHFHYVMSLGA-- 419  
DB 374 GFVAFEFVGPGPIPMFTIVALESGQPRPAIIVAGFSN--WTSNFIYVGMCFQYVDELQGPY 432  
QY 420 IFALFAGI-----YFMPKFSGRAPEWAA 444  
DB 433 VFILFTVLVLFIRTYFKVPETKGRTFDEIAS 465

RESULT 5  
US-09-031-392-4  
Sequence 4, Application US/09031392  
Patent No. 5942398  
GENERAL INFORMATION:  
APPLICANT: Tartaglia, Louis A.  
APPLICANT: Wang, Xun  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street

```

: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Melkijohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 534 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-031-392-4

```

```

Query Match      3.5%; Score 106; DB 2; Length 534;
Best Local Similarity 21.2%; Pred. No. 0.031;
Matches 98; Conservative 58; Mismatches 155; Indels 152; Gaps 22;

```

```

QY 40 VVGFTSVLFTVYVRLMDPGVQVYMCLEGARL-----IADASQTCANGH 84
DB 142 VVNVLSIAGNLMGLAKMPS-HILITAGRAITGLYGLSSGLVPMYSEVSPTA----- 195
QY 85 LNNWVYTH-----GILMFPVGIPLFGGNGNYLMPLOIGAPDMAPRRMNLSEFWLFI 138
DB 196 LRGAIGTLHQIAIVTGILISQVIGLDFLLG--NDELMPILLGLSGVA---ALLQFELL- 248
QY 139 AGTAMGVASLFAFGGDCGLSGVGWVLYPLSTREAGYSMDLAIFAVHLSGASSIMGAIN 198
DB 249 -----LLCPESPXYLYIKLGKV-----EAKKSLK-----RLRGNCDPMKEIA 286
QY 199 MITTELNMARPGMTLHKVPLFESWISFTTAMILLALPYLAGATMLLTDNR---FGTF 254
DB 287 EMEKEKQEAASEKRVISIGQLFSSSKYRQA--VIVALMV---QISQOFGINAIIFYSTNI 341
QY 255 FNPAGGDPILYOHILMFPGHPEVYIIILPFGIISHVSTFS-----KKVPFGYLP 306
DB 342 FQRAVGQPYVYATL-----GCVGVNVTFTVIVSLVEKAGRSLSF----- 382
QY 307 MYTAMVAICVLGFVVAHHMTVGMSTLQOQSYFMLATVIAVPTGIKITSWATMMGGSV 366
DB 383 -----LAGLMGMLISAVAM-TVGLVLLSQ-----FAMMSYV----- 412
QY 367 EFKSPMLAEGFMFLFTVG-----GVTGIYLAQAGLDRAVHDITYVVAHF 411
DB 413 -----SWAIFLFTVFEVGGPIPMFIVALELFSQGPRAAIAVAGFCN-WACNFIVGKCF 467
QY 412 HYVMSLGA--IFAIFAGI-----YFYMPKFSGRAPFEMA 444
DB 468 QYIADLCGPYVVFVAVLLVFLFAVLAFLKVPETKGSFEEIAA 510

```

```

RESULT 6
US-09-299-549-4
: Sequence 4, Application US/09299549
: Patent No. 6136547
: GENERAL INFORMATION:
: APPLICANT: Tartaglia, Louis A.
: APPLICANT: Weng, Xun

```

```

: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
: TITLE OF INVENTION: GLUTEX AND USES THEREOF
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/299,549
: FILING DATE: 26-APR-1999
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Melkijohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 534 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-299-549-4

```

```

Query Match      3.5%; Score 106; DB 4; Length 534;
Best Local Similarity 21.2%; Pred. No. 0.031;
Matches 98; Conservative 58; Mismatches 155; Indels 152; Gaps 22;

```

```

QY 40 VVGFTSVLFTVYVRLMDPGVQVYMCLEGARL-----IADASQTCANGH 84
DB 142 VVNVLSIAGNLMGLAKMPS-HILITAGRAITGLYGLSSGLVPMYSEVSPTA----- 195
QY 85 LNNWVYTH-----GILMFPVGIPLFGGNGNYLMPLOIGAPDMAPRRMNLSEFWLFI 138
DB 196 LRGAIGTLHQIAIVTGILISQVIGLDFLLG--NDELMPILLGLSGVA---ALLQFELL- 248
QY 139 AGTAMGVASLFAFGGDCGLSGVGWVLYPLSTREAGYSMDLAIFAVHLSGASSIMGAIN 198
DB 249 -----LLCPESPXYLYIKLGKV-----EAKKSLK-----RLRGNCDPMKEIA 286
QY 199 MITTELNMARPGMTLHKVPLFESWISFTTAMILLALPYLAGATMLLTDNR---FGTF 254
DB 287 EMEKEKQEAASEKRVISIGQLFSSSKYRQA--VIVALMV---QISQOFGINAIIFYSTNI 341
QY 255 FNPAGGDPILYOHILMFPGHPEVYIIILPFGIISHVSTFS-----KKVPFGYLP 306
DB 342 FQRAVGQPYVYATL-----GCVGVNVTFTVIVSLVEKAGRSLSF----- 382
QY 307 MYTAMVAICVLGFVVAHHMTVGMSTLQOQSYFMLATVIAVPTGIKITSWATMMGGSV 366
DB 383 -----LAGLMGMLISAVAM-TVGLVLLSQ-----FAMMSYV----- 412
QY 367 EFKSPMLAEGFMFLFTVG-----GVTGIYLAQAGLDRAVHDITYVVAHF 411
DB 413 -----SWAIFLFTVFEVGGPIPMFIVALELFSQGPRAAIAVAGFCN-WACNFIVGKCF 467
QY 412 HYVMSLGA--IFAIFAGI-----YFYMPKFSGRAPFEMA 444
DB 468 QYIADLCGPYVVFVAVLLVFLFAVLAFLKVPETKGSFEEIAA 510

```





LENGTH: 617 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Brain -norepinephrine transporter  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 95..96  
OTHER INFORMATION: /note="Leucine zipper motif"  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 102..103  
OTHER INFORMATION: /note="Leucine zipper motif"  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 109..110  
OTHER INFORMATION: /note="Leucine zipper motif"  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 116..117  
OTHER INFORMATION: /note="Leucine zipper motif"  
US-08-753-985-11

Query Match 3.5%; Score 106; DB 1; Length 617;  
Best Local Similarity 19.8%; Pred. No. 0.038;  
Matches 69; Conservative 43; Mismatches 117; Indels 120; Gaps 15;

QY 95 ILMMFVGIPLFGFGNYLMPLOIGAPDMAPRRNNLSFWL-----FIAGTANGVAS 147  
DB 276 VLLHGVTLTPGASNGINAYL-----HIDFYRLKATVWIDATQIFSLGAGFGVLI 327  
QY 148 LFAPGG-----DQGL-----GSGVG 162  
DB 328 AFASYNKFDNNCYRDALLTSSINCITSFVSGFATFSLIGYMAHEKVNIEDVATGAGLV 387  
QY 163 WLYPLPLSTREAGYSMDLAFVAVL--SGASSIMGAINMITTFINMRAPGMLHKVPLFS 220  
DB 388 FILYPEAISTLSGSTFMVAVFVFMALLGLDSSMGMEAVITGLADDPVYLRHR-KLFT 446  
QY 221 WSIFITALLILALPYLAGATMLLTDRNFGTFFNPAGGDPILYQHIL-----WFG 274  
DB 447 FGVTFSFTLLALFCITKGGIYVLTLD-----TF-----AAGTSILFAVLMGAGVSWFYG 497  
QY 275 -----HPEVY-----IILPGFIISHVVSFSSKKPVF--GYLPMVYAMVA 313  
DB 498 VDRFSNDIQOMGFRPGLIYWLCKKFPVSPALLLVVVVSIINFKPLTYDDYI----- 549  
QY 314 IGVIGFVVAHMHMYTVGMSLTQOSYFMLATMVAVPTGIKIFSMTATMW 362  
DB 550 -----FPPMAN--WVGWGIALLSMVLPVIYI-----YKFLSTQSGSLW 585

RESULT 11  
US-09-084-813-3  
Sequence 3, Application US/09084813  
Patent No. 6127131  
GENERAL INFORMATION:  
APPLICANT: Smith, Kelli  
APPLICANT: Borden, Laurence A.  
APPLICANT: Branchek, Theresa  
APPLICANT: Hartig, Paul R.  
APPLICANT: Weishank, Richard L.  
TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/084,813  
FILING DATE: 26-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1795/39875-AZ-PCIT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 617 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: N  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: HUMAN NORADRENALINE TRANSPORTER  
US-09-084-813-3

Query Match 3.5%; Score 106; DB 3; Length 617;  
Best Local Similarity 19.8%; Pred. No. 0.038;  
Matches 69; Conservative 43; Mismatches 117; Indels 120; Gaps 15;

QY 95 ILMMFVGIPLFGFGNYLMPLOIGAPDMAPRRNNLSFWL-----FIAGTANGVAS 147  
DB 276 VLLHGVTLTPGASNGINAYL-----HIDFYRLKATVWIDATQIFSLGAGFGVLI 327  
QY 148 LFAPGG-----DQGL-----GSGVG 162  
DB 328 AFASYNKFDNNCYRDALLTSSINCITSFVSGFATFSLIGYMAHEKVNIEDVATGAGLV 387  
QY 163 WLYPLPLSTREAGYSMDLAFVAVL--SGASSIMGAINMITTFINMRAPGMLHKVPLFS 220  
DB 388 FILYPEAISTLSGSTFMVAVFVFMALLGLDSSMGMEAVITGLADDPVYLRHR-KLFT 446  
QY 221 WSIFITALLILALPYLAGATMLLTDRNFGTFFNPAGGDPILYQHIL-----WFG 274  
DB 447 FGVTFSFTLLALFCITKGGIYVLTLD-----TF-----AAGTSILFAVLMGAGVSWFYG 497  
QY 275 -----HPEVY-----IILPGFIISHVVSFSSKKPVF--GYLPMVYAMVA 313  
DB 498 VDRFSNDIQOMGFRPGLIYWLCKKFPVSPALLLVVVVSIINFKPLTYDDYI----- 549  
QY 314 IGVIGFVVAHMHMYTVGMSLTQOSYFMLATMVAVPTGIKIFSMTATMW 362  
DB 550 -----FPPMAN--WVGWGIALLSMVLPVIYI-----YKFLSTQSGSLW 585

RESULT 12  
PCT-US92-09662-3  
Sequence 3, Application PC/TUS9209662  
GENERAL INFORMATION:  
APPLICANT: Smith, Kelli  
APPLICANT: Borden, Laurence A.  
APPLICANT: Branchek, Theresa  
APPLICANT: Hartig, Paul R.  
APPLICANT: Weishank, Richard L.



RESULT 14  
US-09-031-392-3  
; Sequence 3, Application US/09031392  
; Patent No. 5942398  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Meng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/031.392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 383 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-031-392-3

Query Match 3.3%; Score 99.5; DB 2; Length 383;  
Best Local Similarity 20.8%; Pred. No. 0.087;  
Matches 81; Conservative 42; Mismatches 134; Indels 133; Gaps 16;

QY 137 FLAGMAGVASLFPAGGCGOLSGVGWVLYPLSTREA-----GYSMDLAIFAHLGSGS 191  
DB 18 FLIGVYCGLTGTFVPMYGEVS-----PTELKGLGTLHQLGIVGILIAOVFGLD 68  
QY 192 SIMGAINMITTFELNMRAPGMLHKVPLFMSJFTTAMLILALPVLAGATMLLTDRN-- 249  
DB 69 SINGMDEL-----WPLLSTVIFIPALLOCLILFPCBPSPRLINRNEE 112  
QY 250 -----FGTT-----FFNPAGGDPILYOHILMF 273  
DB 113 NRAKSVLKLKRGADVTDRDLOEMKEESROMREKVTLLLEFRSAAYROPILIAVVLQS 172  
QY 274 GHEVEYIIILPGFGIISHVVSFESKRPVFGYLPVYVYMAVIGVLS-----FVY---- 321  
DB 173 QO-----LSGINAVFYSTISFEKA--GVQDPYATIGSGIVNTAFTVVSLEFVERAG 223  
QY 322 --WAHMYTVGNSLTOOSYFMLATVIAVPTGIRFISWATMTMGGSVEFKSPMLAEGFM 379  
DB 224 RRTLHLIGLAGMA-----GCAVLMITLAL--LEQLPMSYLS-----SIVAIIFGV 267  
QY 380 FLFTVG-----GVTGIYLAQAGLDRAVHDITYVVAHFHYVMSLGA--IFA 422  
DB 268 AFREVEGPPIPMFIVALEFSQGRPAIAVAGFSN--WTSNFTVMCFQYVEQLCGPYFI 326  
QY 423 IFAGI-----YFMPKESGRAPPEMAA 444  
DB 327 IFTVLVLEFIFTFYRKVPETKGTDEIAS 356

RESULT 15  
US-09-299-549-3  
; Sequence 3, Application US/09299549  
; Patent No. 6136547  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Meng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299.549  
; FILING DATE: 26-APR-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/031.392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 383 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-299-549-3

Query Match 3.3%; Score 99.5; DB 4; Length 383;  
Best Local Similarity 20.8%; Pred. No. 0.087;  
Matches 81; Conservative 42; Mismatches 134; Indels 133; Gaps 16;

QY 137 FLAGMAGVASLFPAGGCGOLSGVGWVLYPLSTREA-----GYSMDLAIFAHLGSGS 191  
DB 18 FLIGVYCGLTGTFVPMYGEVS-----PTELKGLGTLHQLGIVGILIAOVFGLD 68  
QY 192 SIMGAINMITTFELNMRAPGMLHKVPLFMSJFTTAMLILALPVLAGATMLLTDRN-- 249  
DB 69 SINGMDEL-----WPLLSTVIFIPALLOCLILFPCBPSPRLINRNEE 112  
QY 250 -----FGTT-----FFNPAGGDPILYOHILMF 273  
DB 113 NRAKSVLKLKRGADVTDRDLOEMKEESROMREKVTLLLEFRSAAYROPILIAVVLQS 172  
QY 274 GHEVEYIIILPGFGIISHVVSFESKRPVFGYLPVYVYMAVIGVLS-----FVY---- 321  
DB 173 QO-----LSGINAVFYSTISFEKA--GVQDPYATIGSGIVNTAFTVVSLEFVERAG 223  
QY 322 --WAHMYTVGNSLTOOSYFMLATVIAVPTGIRFISWATMTMGGSVEFKSPMLAEGFM 379  
DB 224 RRTLHLIGLAGMA-----GCAVLMITLAL--LEQLPMSYLS-----SIVAIIFGV 267  
QY 380 FLFTVG-----GVTGIYLAQAGLDRAVHDITYVVAHFHYVMSLGA--IFA 422

Db 268 AFEEVGPPIPMFIVAEIIFSGPRPAAIAVAGEFN-WTSNFIWGMCFQYVEOLGPPVFI 326  
Qy 423 IFAGI-----YFYMPKFSGRAPEMAA 444  
|| : || : || : || :  
Db 327 IFTVLLVLFIFTYFKVPETKGRFDEIAS 356

Search completed: December 22, 2001, 11:02:58  
Job time: 337 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 10:55:21 ; Search time 24.27 Seconds  
(without alignments)  
1748.217 Million cell updates/sec

Title: US-09-712-768-2

Perfect score: 3000

Sequence: 1 MADAAIGHDHHEKGGFFTR.....PAHTEETLPKRSMDKHPSH 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2538.5	84.6	558	2 S08270	cytochrome-c oxida
2	2453	81.8	565	2 S20534	cytochrome-c oxida
3	2347	78.2	554	2 S03809	cytochrome-c oxida
4	1873.5	62.5	538	2 S36424	cytochrome-c oxida
5	1829	61.0	541	1 ODZ01	cytochrome-c oxida
6	1773	59.1	523	2 S62706	cytochrome-c oxida
7	1765.5	58.5	522	2 S25956	cytochrome-c oxida
8	1748.5	58.3	524	2 S14138	cytochrome-c oxida
9	1745	58.2	515	2 T11913	cytochrome-c oxida
10	1744.5	58.1	527	2 S65346	cytochrome-c oxida
11	1744	58.1	534	2 D71698	cytochrome-c oxida
12	1738.5	58.0	527	2 S14139	cytochrome-c oxida
13	1730.5	57.7	524	1 ODR21	cytochrome-c oxida
14	1730.5	57.7	524	1 ODR21	cytochrome-c oxida
15	1729.5	57.6	524	2 S16256	cytochrome-c oxida
16	1713.5	57.1	514	2 S30281	cytochrome-c oxida
17	1707	56.9	537	2 T11237	cytochrome-c oxida
18	1698.5	56.6	531	2 S78190	cytochrome-c oxida
19	1677.5	55.9	532	2 S59087	cytochrome-c oxida
20	1677	55.9	536	2 S63651	cytochrome-c oxida
21	1670	55.7	533	2 A58931	cytochrome-c oxida
22	1651.5	55.0	527	1 OBSY1	cytochrome-c oxida
23	1643.5	54.8	527	2 S05290	cytochrome-c oxida
24	1642	54.7	531	2 T112406	cytochrome-c oxida
25	1639	54.6	530	2 T11884	cytochrome-c oxida
26	1609	53.6	557	1 ODNC1	cytochrome-c oxida
27	1608.5	53.6	527	1 ODOB1M	cytochrome-c oxida
28	1605.5	53.4	567	1 ODA51	cytochrome-c oxida
29	1600.5	53.4	471	2 T43827	cytochrome-c oxida

30	1600	53.3	528	2 S26948	cytochrome-c oxida
31	1594	53.1	526	2 S62763	cytochrome-c oxida
32	1594	53.1	541	2 A48327	cytochrome-c oxida
33	1588.5	53.0	517	2 C34284	cytochrome-c oxida
34	1564.5	52.1	517	2 S01501	cytochrome-c oxida
35	1562.5	52.1	517	2 S14205	cytochrome-c oxida
36	1561.5	52.0	873	1 S53828	cytochrome-c oxida
37	1561	52.0	519	2 T11129	cytochrome-c oxida
38	1558.5	52.0	517	2 S70597	cytochrome-c oxida
39	1558	51.9	516	2 T11182	cytochrome-c oxida
40	1554.5	51.8	512	1 ODBT1	cytochrome-c oxida
41	1552	51.7	516	2 T11195	cytochrome-c oxida
42	1550	51.7	513	2 T11482	cytochrome-c oxida
43	1548.5	51.6	534	2 S78640	cytochrome-c oxida
44	1546	51.5	516	2 S45351	cytochrome-c oxida
45	1545.5	51.5	534	2 JU0148	cytochrome-c oxida

#### ALIGNMENTS

##### RESULT 1

S08270 cytochrome-c oxidase (EC 1.9.3.1) chain I-beta - Paracoccus denitrificans

C:Species: Paracoccus denitrificans

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-May-2000

C:Accession: S08270; C35121

R:Raitto, M.; Pispas, J.M.; Metsu, T.; Saraste, M.

FEBS Lett. 261, 431-435, 1990

A:Title: Are there isoenzymes of cytochrome c oxidase in Paracoccus denitrificans?

A:Reference number: S08269; M01D:90184495

A:Accession: S08270

A:Molecule type: DNA

A:Residues: 1-558 <RAT>

A:Cross-references: EMBL:Y07533; NID:g45477; PIDN:CAA68821.1; PID:g45479

R:Van Spanning, R.J.M.; Mansell, C.; Harms, N.; Oltmann, L.F.; Stoutamer, A.H.

J. Bacteriol. 172, 986-996, 1990

A:Title: Mutagenesis of the gene encoding cytochrome c-550 of Paracoccus denitrifican

A:Reference number: A35121; M01D:90130336

A:Accession: C35121

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-58 <VAN>

A:Cross-references: GB:M27304; NID:g150573; PIDN:AA88365.1; PID:g150575

R:Busse, G.; Soullame, T.; Dewor, M.; Meyer, H.E.; Bluggel, M.

Protein Sci. 8, 985-990, 1999

A:Title: Evidence for a copper-coordinated histidine-tyrosine cross-link in the activ

A:Reference number: A58960; M01D:99268331

A:Contents: annotation

A:Note: mass spectrophagic and chemical characterization of histidyl-tyrosine cross-

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: Chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane

F:27-95/Domain: cytochrome-c oxidase chain I homology <C01>

F:27-95/Domain: transmembrane #status experimental <TM01>

F:84-121/Domain: transmembrane #status experimental <TM02>

F:130-151/Domain: transmembrane #status experimental <TM03>

F:178-206/Domain: transmembrane #status experimental <TM04>

F:218-251/Domain: transmembrane #status experimental <TM05>

F:263-298/Domain: transmembrane #status experimental <TM06>

F:304-322/Domain: transmembrane #status experimental <TM07>

F:334-362/Domain: transmembrane #status experimental <TM08>

F:370-395/Domain: transmembrane #status experimental <TM09>

F:404-430/Domain: transmembrane #status experimental <TM10>

F:441-468/Domain: transmembrane #status experimental <TM11>

F:483-513/Domain: transmembrane #status experimental <TM12>

F:66-80/Disulfide Bonds: #status experimental

F:94,413/Binding site: heme a iron (His) (axial ligands) #status experimental

F:276-280/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status experimental

F:280/Binding site: oxygen (Tyr) #status predicted

F:403/Binding site: magnesium (His) (shared with chain II) #status predicted

F:411/Binding site: heme a3 iron (His) (axial ligand) #status experimental

Query Match 84.6%; Score 2538.5; DB 2; Length 558;  
Best Local Similarity 81.0%; Pred. No. Se-169;  
Matches 452; Conservative 51; Mismatches 54; Indels 1; Gaps 1;

OY 1 MADAAIHGH-DHHEKOGFTRMENSTNKHDKIGLLYLVAAGVGFSTVFYMRLEMDP 59  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 1 MADAAVHHGHHDRGGFTTRFMSTNHKDGILXLFPAAGIVGLISVCFTYYMMMELOHP 60  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 60 GVQYMCLEGARLIADASOTCTANGHLNMNVNTYHGIIIMFEPVGIPALFGSGNYLPLQI 119  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 61 GVQYMCLEGARLIADASAECPTRNGHLNMVTITYGVLMMFPVVIPALEGGGNFMPLEHI 120  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 120 GAPDMAFRPMNNLSWFLFIAGTAMGVASLAFAPGDGOLGSQVWVLPRLSTRAGISMD 179  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 121 GAPDMAFRPLNLTSLMTMVCGVALGVASLAFPGNDGSSGVGWVLPRLSTTAGYSMD 180  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 180 LAIFAVHLSGASSINGAINTMTPFLNMRPQMTHLKVPFLFSMSIFTAWILLALPYLAG 239  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 181 LAIFAVHVSAGASSILGAINTMTPFLNMRPQMTHLKVPFLFMSVFTAWILLALSPLYLAG 240  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 240 AITMLLTRNFEGTFEFPNAGGDPRILOHILMFEGHPENVYIIILPGFIIISHVSTESKK 299  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 241 AITMLLMRNETGTQDFPDAGGDDPVLYQHILMFEGHPENVYIIILPGFIIISHVSTFAKK 300  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 300 PVFGLPLPWYANVAIGVLCFYVMANHMTYVGMNSTQDSYFMLATVNAVPRGIKIFSIA 359  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 301 PLFGYLPWYLAAMAAGIIGFYVMANHMYTAQNSILOQAIFYMLATMTAVNPVGIKVSIFA 360  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 360 TWMGSSVEFKSPMLAFQFMFLFTVGVTGYILAQAGIDRAYHDITYVVAHFHYMSIGA 419  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 361 TWMGSIIEFKTMIAFGELFLFTYGVGTGYVLSAPLRDXYHDIYYVAHFHYMSIGA 420  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 420 IFAIFAGIIFYMPKFSGRAFPDEMAKLHFWTFETIGANTFEPPHFLRGQMPRRYIDYE 479  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 421 VEGIFAGYVYMVKSGSQRYREMAQLHFHMWFIGSNLIFPPQHLRGQMPRRYIDYV 480  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 480 AFALNNKSVSGAFLAFAASFLEFYTYTLVAGRERTRPNMGFADTLWTLPSPPPA 539  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 481 EFAYWNNTSISAIYSPASFLEFYTYTLVAGRKNVNPVNMEDHADTLWTLPSPPE 540  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
OY 540 HFEETLPKRSDWDKHPSH 557  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 541 HFEETLPKRSDMDRAHAH 558  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2  
S20534  
cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhodobacter sphaeroides  
C:Species: Rhodobacter sphaeroides  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text-change 20-Aug-1999  
C:Accession: S20534; A42378  
R:Shapleigh, J.P.; Gibbs, R.B.  
Mol. Microbiol. 6, 635-642, 1992  
A>Title: Cloning, sequencing and deletion from the chromosome of the gene encoding subunit  
A:Reference number: S20534; MUID:92204019  
A:Accession: S20534  
A:Molecule type: DNA  
A:Residues: 1-565 <SHA>  
A:Cross-references: EMBL:X62645; NTD:946417; PIDN:CAAA4514.1; PID:946418  
A:Note: the authors translated the codon AAC for residue 214 as Met and ATG for residue  
R:Shapleigh, J.P.; Hill, J.O.; Alden, J.O.; Gibbs, R.B.  
J. Bacteriol. 174, 2338-2343, 1992  
A>Title: Spectroscopic and genetic evidence for two heme-Cu-containing oxidases in Rhodospirillum rubrum  
A:Reference number: A42378; MUID:92202164  
A:Accession: A42378  
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 286-289, 'L', 292-293, 'G', 295-296, 'M', 299, 'QIVS', 304-305, 'SR', 308-309, 'V', 310-  
311, 'AARLN', 'S', 486, 'S', 488, 'S' <SH2>  
C:Genetics:  
A:Gene: ctad  
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: cytochrome; copper; electron transfer; heme; iron; magnesium; membrane  
F:25-500/Domain: cytochrome-c oxidase chain I homology <COI>  
F:102,421/Binding site: heme a iron (His) (axial ligands) #status predicted  
F:284,333,334/Binding site: copper (His) #status predicted  
F:284-288/Cross-link: 1'-histidyl-3'-lysine (His-Tyr) #status predicted  
F:288/Binding site: oxygen (Tyr) #status predicted  
F:411/Binding site: magnesium (His) (shared with chain II) #status predicted  
F:419/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match . 81.8%; Score 2453; DB 2; Length 565;  
Best Local Similarity 78.7%; Pred. No. 4,4e-163;  
Matches 446; Conservative 50; Mismatches 59; Indels 12; Gaps 3;

Db 1 MADAAIHGDHHEKQGFTRFMSTNKHDDGLLYVAAGVGFISLFTYVMRELMDPG 60  
| :  
1 MADAAIHGHE -HDRGRFTFRMFSNKHDDGLYLFTGGDLGLLSVAFETYMRRLMAPG 59  
QY 61 VOYMLEGAR-----LTADASOTCTANGHLAMNVVHYGHLLMFEEFGIALPGEF 110  
| : : : | :  
60 VDFPCAHELESIVGFFQSILMPASAVENCTENGHLAMNVITGHGLIMFEVITALEGGF 119  
Db 111 GNLYMLPLQIGADMAFPRRNNLSFWLFIATGAMGVASLFAPGSGDGLSGVGWLVLPPLS 170  
| :  
120 GNYFPLRHIGADMAFPRRNNLSYVLVAAGSLVASLFAFGNGOGLSGGWLVLPPLS 179  
Db 171 TREAGYSMDIAFAVHLSGASSIMGALMITTFELNMRAFGMTLHKVPLEFSIFTTAML 230  
| :  
180 TSEGSTDLDAFAVHLSGASSILGAILMITYTEFLMKRPAFGMTMKRVLPFASITVTAML 239  
QY 231 LLAELVLGAITMLTLDSNFGTTFPNPAGGDDLTYOHILMFEGHREUYIIIRPGGIIS 290  
| :  
240 LLALPVLAGAITMLTLDRNEGTTFPGSGGGDPVLYCHILMFEGHREUYIILPAFGIVS 299  
QY 291 HVVSFESKKRPVCGYLPWYAMAIVLGFWYAHNMITYVGSLSLQQSYFMLATWIAVPT 350  
| :  
300 HVIATFAFKRIEGLPYMYAMAIVLGFWYAHNMITYVGSLSLQQSYFMLATWIAVPT 359  
Db 351 GKIFTSWITAMGSGVEERKSPMLAFCFMLEFTVGVGCVILAQGLDRAVHDYVVVAH 410  
| :  
360 GKIFTSWITAMGSSIELKTPMLNALGLELFTVGVGCVIYLSQASVDRYHHDIYYVAH 419  
QY 411 FHYVMSLGAIEAIFAGIFYPMKFSGRAFPEWAAKLHFWEFIGANTVFEPONFLRGOM 470  
| :  
420 FHYVMSLGAIVEGIFAGISGIGKMSGRQYPEAGKLHFMFVGNLFFPOHFLRGOM 479  
Db 471 PRRYIDYPAAFPLMKVSSYGAFLAASFLEPIFYEVTVIAGREPFRNPGEADPLE 530  
| :  
480 PRRYIDYPAAFPLMVFVSSLGAFLSFASFLPGLGYFSL-SGARVANANNWNEDADTLE 538  
QY 531 WTLPSPPAHPEETLPKRSDWDKHPHS 557  
| :  
Db 539 WTLSPRPHTPEQLPKREDMERAPAH 565

RESULT 3  
S03809  
cytochrome-c oxidase (EC 1.9.3.1) chain I - Paracoccus denitrificans  
C:Species: Paracoccus denitrificans  
C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 20-Jun-2000  
C:Accession: S03809  
R:Raitio, M.; Jalli, T.; Saraste, M.  
EMBO J. 6, 2825-2833, 1987  
A>Title: Isolation and analysis of the genes for cytochrome c oxidase in Paracoccus d  
A:Reference number: S03803  
A:Accession: S03809  
A:Molecule type: DNA  
A:Residues: 1-554 <RAI>  
A:Cross-references: EMBL:X05829; NID:994806; PIDN:CAA29274.1; PID:9995678  
C:Genetics:  
A:Gene: COI  
C:Keywords: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane

F:24-487/Domain: cytochrome-c oxidase chain I homology <COI>  
 F:91,408/Binding site: heme a iron (His) (axial ligand) #status predicted  
 F:273,322,333/Binding site: copper (His) #status predicted  
 F:273-277/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:277/Binding site: oxygen (Tyr) #status predicted  
 F:396/Binding site: magnesium (His) (shared with chain II) #status predicted  
 F:406/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 78.2% Score 2347; DB 2; Length 554;

Best Local Similarity 76.1% Pred. No. 9.9e-156; Matches 415; Conservative 59; Mismatches 69; Indels 2; Gaps 1;

QY 13 EKQGFETRMSTNKHDKGLYLVAAGVGFISVLEFYVMELEMDPQVYMCLEGARLI 72  
 DB 11 EKRGFETRMSTNKHDKGLYLVAAGVGFISVLEFYVMELEMDPQVYMCLEGARLI 70  
 QY 73 ADASOTCTANGHNLNVMVVTYHGLIMFVFGIPALFGFGNYLMPLOIGAPDAPRRMNL 132  
 DB 71 ADAAECPNMAHNLNVMVVTYHGLIMFVFGIPALFGFGNYLMPLOIGAPDAPRRMNL 130  
 QY 133 SFWLEIAGTANGVSLFAPGDDGOLGSGGVNLYPPLSTRAGYSMDLAFVHLGSSS 192  
 DB 131 SYWLVCVSLAIVSLSPGSDOPGAGVNVLYPPLSTRAGYAMDALFAVHVSQATSS 190  
 QY 193 IMGAINMTTFLNMRAPGMTLHKVPLFESMIFITAMILLALPVLACATMLLTDNRFGT 252  
 DB 191 ILGAINITTTFLNMRAPGMTLHKVPLFAMAVFITAMILLALPVLACATMLLTDNRFGT 250  
 QY 253 TFEFNAGGDDPLVOHILMFQHPREVYIILPGGITSHVVSFESKRPVEGYLPMYAV 312  
 DB 251 OFEDPAGGDDPLVOHILMFQHPREVYIILPGGITSHVVSFESKRPVEGYLPMYAV 310  
 QY 313 AIGVGFVMAHMYTVGMSLTQOSTFEMLATMVAIVPTGIFESLAIMMGSGVEKSPM 372  
 DB 311 AIAFGFTVMAHMYTVGMSLTQOSTFEMLATMVAIVPTGIFESLAIMMGSGVEKSPM 370  
 QY 373 LMAFGFMLEFTVGGVTVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAFAIGIYEMP 432  
 DB 371 LMAFLA-LEFTVGGVTVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAFAIGIYEMP 428  
 QY 433 KESGAFDEMAAKLHFMTFFIGANVTFFPHFLRGHGMRRYIDYDEPAFALMKNVSSYGA 492  
 DB 429 KMSGQVDEMAAKLHFMTFFIGANVTFFPHFLRGHGMRRYIDYDEPAFALMKNVSSYGA 488  
 QY 493 FLAFASTFLFYVETVLVAGRRETRPNMGEPADTLEMTLPSPPAHTFETLKRSDMD 552  
 DB 489 YISFASSTFLFYVETVLVAGRRETRPNMGEPADTLEMTLPSPPAHTFETLKRSDMD 548  
 QY 553 KHPSH 557  
 DB 549 RAQAH 553

RESULT 4  
 S36424  
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhizobium leguminosarum

C:Species: Rhizobium leguminosarum  
 C:Date: 09-Dec-1993 #sequence\_revision 03-Nov-1995 #text\_change 23-Jul-1999  
 C:Accession: S36424  
 R:Gabel, C.; Bittlinger, M.A.; Maier, R.J.  
 submitted to the EMBL Data Library, July 1993  
 A:Description: Cytochrome aa3 gene regulation in the Rhizobiaceae: comparison of copper  
 A:Reference number: S36424  
 A:Accession: S36424  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-538 <GAB>  
 A:Cross-references: EMBL:X74341; NID:g396698; PIDN:CAA52388.1; PID:g396699  
 C:Genetics:  
 A:Gene: coxa  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as

rotein  
 F:31-481/Domain: cytochrome-c oxidase chain I homology <COI>  
 F:82,402/Binding site: heme a iron (His) (axial ligand) #status predicted  
 F:255,314,315/Binding site: copper (His) #status predicted  
 F:265-269/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:269/Binding site: oxygen (Tyr) #status predicted  
 F:392/Binding site: magnesium (His) (shared with chain II) #status predicted  
 F:400/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 62.5% Score 1873.5; DB 2; Length 538;

Best Local Similarity 64.2% Pred. No. 7.2e-123; Matches 348; Conservative 63; Mismatches 110; Indels 21; Gaps 4;

QY 7 HGHDEHAKQGFETRMSTNKHDKGLYLVAAGVGFISVLEFYVMELEMDPQVYMCLE 66  
 DB 15 HGHDEHAKQGFETRMSTNKHDKGLYLVAAGVGFISVLEFYVMELEMDPQVYMCLE 69  
 QY 67 EGARLIADASOTCTANGHNLNVMVVTYHGLIMFVFGIPALFGFGNYLMPLOIGAPDMAF 126  
 DB 70 -----HNTLNLNVMVVTYHGLIMFVFGIPALFGFGNYLMPLOIGAPDMAF 115  
 QY 127 PRMNLSFWLPLAGTANGVSLFAPGDDGOLGSGGVNLYPPLSTR-EGAGSMDLAFV 185  
 DB 116 PRMNLSFWLPLAGTANGVSLFAPGDDGOLGSGGVNLYPPLSTR-EGAGSMDLAFV 175  
 QY 186 HLGASSIMGAINMTTFLNMRAPGMTLHKVPLFESMIFITAMILLALPVLACATMLL 245  
 DB 176 HLGASSIMGAINMTTFLNMRAPGMTLHKVPLFESMIFITAMILLALPVLACATMLL 235  
 QY 246 TDRNFGTTFEPNAGGDDPLVOHILMFQHPREVYIILPGGITSHVVSFESKRPVEGYL 305  
 DB 236 TDRNFGTTFEPNAGGDDPLVOHILMFQHPREVYIILPGGITSHVVSFESKRPVEGYL 295  
 QY 306 RMYAVMAIVGLFVVAHMYTVGMSLTQOSTFEMLATMVAIVPTGIFESLAIMMGSG 365  
 DB 296 RMYAVMAIVGLFVVAHMYTVGMSLTQOSTFEMLATMVAIVPTGIFESLAIMMGSG 355  
 QY 366 VEPKSPMLAEGFLEFTVGGVTVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFA 425  
 DB 356 VEPKSPMLAEGFLEFTVGGVTVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFA 415  
 QY 426 GIFYMFKFGRAFPREMAAKLHFMTFFIGANVTFFPHFLRGHGMRRYIDYDEPAFALM 485  
 DB 416 GIFYMFKFGRAFPREMAAKLHFMTFFIGANVTFFPHFLRGHGMRRYIDYDEPAFALM 475  
 QY 486 KVSYGAFLEFASLFTFYVTVLVAAGRRETRPNMGEPADTLEMTLPSPPAHTFETL 545  
 DB 476 LVSSIGSYISGFSVLEFYIC-VYDAFAKNVPVGDNDGAGATTLLEMTLPSPPVHEFEVL 534  
 QY 546 PK 547  
 DB 535 PR 536

RESULT 5  
 ODZJ1  
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Bradyrhizobium japonicum

C:Species: Bradyrhizobium japonicum  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 23-Jul-1999  
 C:Accession: S13076; S12101  
 R:Botl, M.; Bolliger, M.; Hennecke, H.  
 Mol. Microbiol. 4, 2147-2157, 1990  
 A:Title: Genetic analysis of the cytochrome c-aa(3) branch of the Bradyrhizobium japo  
 A:Reference number: S13076; M01D:91211625  
 A:Accession: S13076  
 A:Molecule type: DNA  
 A:Residues: 1-541 <BOT>  
 A:Cross-references: EMBL:X54800; NID:g39502; PIDN:CAA38570.1; PID:g39503  
 R:Gabel, C.; Maier, R.J.  
 Nucleic Acids Res. 18, 6143, 1990  
 A:Title: Nucleotide sequence of the coxa gene encoding subunit I of cytochrome aa(3)  
 A:Reference number: S12101; M01D:91045095

A:Accession: S12101  
 A:Molecule type: DNA  
 A:Residues: 1-541 <GAB>  
 A:Cross-references: EMBL:K54318; NID:g39505; PIDN:CAA3821.1; PID:g39506  
 C:Genetics: coxa  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as  
 rotein  
 F:34-484/Domain: cytochrome-c oxidase chain I homology <CO1>  
 F:41-60/Domain: transmembrane #status predicted <TM01>  
 F:67-107/Domain: transmembrane #status predicted <TM02>  
 F:124-142/Domain: transmembrane #status predicted <TM03>  
 F:173-193/Domain: transmembrane #status predicted <TM04>  
 F:211-231/Domain: transmembrane #status predicted <TM05>  
 F:262-282/Domain: transmembrane #status predicted <TM06>  
 F:294-315/Domain: transmembrane #status predicted <TM07>  
 F:331-352/Domain: transmembrane #status predicted <TM08>  
 F:365-385/Domain: transmembrane #status predicted <TM09>  
 F:404-424/Domain: transmembrane #status predicted <TM10>  
 F:441-460/Domain: transmembrane #status predicted <TM11>  
 F:483-503/Domain: transmembrane #status predicted <TM12>  
 F:545-605/Binding site: heme a iron (His) (axial ligand) #status predicted  
 F:568-617/Binding site: copper (His) #status predicted  
 F:668-722/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:772/Binding site: oxygen (Tyr) #status predicted  
 F:403/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 61.0%; Score 1829; DB 1; Length 541;  
 Best Local Similarity 61.8%; Pred. No. 8,8e-120;

Matches 337; Conservative 72; Mismatches 112; Indels 24; Gaps 4;

7 HGDHHEKQGFPRMFTSTNKHDKIGLLYLVAAGVFISLVFVVRLEMDPGVQYMC 66  
 15 HGDHNAHPGMRKYVSTNKHDKIGLLYLVAAGVFISLVFVVRLEMDPGVQYMC 72  
 67 EGARLADASQSTANGHLLNNVYTHGILMFVFGIPALFGSGFVNLPLQIGADMAF 126  
 73 -----HEHHTVNFVTHGLIMIFVFMVPMAGFGWVPLMIGADMAF 118  
 127 PRNNNLSFVLFTIGTAMGVASLEAPGDDGLSGVGVWLYPRSTR-EGYSDLAIFAV 185  
 119 PRNNNLSFVLFTIGTAMGVASLEAPGDDGLSGVGVWLYPRSTR-EGYSDLAIFAV 178  
 186 HLGASSIMCAIMVITFLPUMRAPGKTLHKVPLFSISITAVLLILALPVLAGATMML 245  
 179 HLGASSIIGAINPITIFPMRAPGKTLHKVPLFSISITAVLLILALPVLAGATMML 238  
 246 TDRNFGTTFPAPGGDPILYOHILMFEGHPEVYIIILPGFIIISVSTFSKPPVGYL 305  
 239 TDRNFGTTFPAPGGDPILYOHILMFEGHPEVYIIILPGFIIISVSTFSKPPVGYL 298  
 306 PMYVAVATGVLSPVVAHMHVTVGMSLQOQSEMLATVIANPFCIKIPSWATMGGS 365  
 299 GMAYAVATGVLSPVVAHMHVTVGMSLQOQSEMLATVIANPFCIKIPSWATMGGS 358  
 366 VEKSPMLAFGFMFLFTVGGVGTIVLAQAGLDRAHNTYVVAHRYVMSLGAIFA 425  
 359 IERRAPIMAVGFIPLFTVGGVGTIVLAQAGLDRAHNTYVVAHRYVMSLGAIFA 418  
 426 GIVYVKEFSGARPEWAKLHMTFTIGANTFPQHLGROGMRVYIDPEAFALNN 485  
 419 GMYVYVKEFSGARPEWAKLHMTFTIGANTFPQHLGROGMRVYIDPEAFALNN 478  
 486 KVSSTGAFALFASFLFFIVYVTLVAGRETRP---NPMGFAADLLENTLSPPPAHNF 542  
 479 LVSSVSYSIG---FGVLLIFLCVDAFAKAVPAGDNPWGAGATLLENTLSPPPHOF 534  
 543 ETLPR 547  
 535 EVLPR 539

# RESULT 6

S62706

Cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) ml

C:Species: mitochondrion Platymonas subcordiformis

C:Date: 19-Mar-1997 #sequence, revision 19-Mar-1997 #text-change 23-Jul-1999

C:Accession: S62706

R:Kessler, U.; Zetsche, K. Plant Mol. Biol. 29, 1081-1086, 1995

A:Title: Physical map and gene organization of the mitochondrial genome from the unic

A:Reference number: S62702; MUID:56145517

A:Accession: S62706

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-523 <KES>

A:Cross-references: EMBL:Z47795; NID:6533584; PIDN:CAA87753.1; PID:6533589

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995

C:Genetics: cox

A:Gene: cox

A:Genome: mitochondrion

C:Function: A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molec

ns from the mitochondrial matrix producing two molecules of water and lowering the co

A:Pathway: oxidative phosphorylation; respiratory chain

A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnes

active phosphorylation; oxidoreductase; respiratory chain; transmembrane protein

F:13-460/Domain: cytochrome-c oxidase chain I homology <CO1>

F:14-42/Domain: transmembrane #status predicted <TM01>

F:14-54/Domain: intracristal #status predicted <IT01>

F:55-90/Domain: transmembrane #status predicted <TM02>

F:91-98/Domain: mitochondrion matrix #status predicted <MM2>

F:99-121/Domain: transmembrane #status predicted <TM03>

F:122-144/Domain: intracristal #status predicted <IT02>

F:145-174/Domain: transmembrane #status predicted <TM04>

F:175-186/Domain: mitochondrion matrix #status predicted <MM3>

F:187-216/Domain: transmembrane #status predicted <TM05>

F:217-231/Domain: intracristal #status predicted <IT03>

F:233-265/Domain: transmembrane #status predicted <TM06>

F:266-322/Domain: mitochondrion matrix #status predicted <MM4>

F:273-288/Domain: transmembrane #status predicted <TM07>

F:290-301/Domain: intracristal #status predicted <IT04>

F:302-330/Domain: transmembrane #status predicted <TM08>

F:331-338/Domain: mitochondrion matrix #status predicted <MM5>

F:339-359/Domain: transmembrane #status predicted <TM09>

F:360-373/Domain: intracristal #status predicted <IT05>

F:374-403/Domain: transmembrane #status predicted <TM10>

F:404-409/Domain: mitochondrion matrix #status predicted <MM6>

F:410-436/Domain: transmembrane #status predicted <TM11>

F:437-449/Domain: intracristal #status predicted <IT06>

F:450-481/Domain: transmembrane #status predicted <TM12>

F:482-522/Domain: mitochondrion matrix #status predicted <MM7>

F:523-581/Binding site: heme a iron (His) (axial ligand) #status predicted

F:582-629/Binding site: copper (His) #status predicted

F:630-648/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted

F:649-678/Binding site: oxygen (Tyr) #status predicted

F:679-732/Binding site: myristate (Lys) (covalent) #status predicted

F:733-737/Binding site: magnesium (His) (shared with chain II) #status predicted

F:738-739/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 59.1%; Score 1773; DB 2; Length 523;  
 Best Local Similarity 60.1%; Pred. No. 6,6e-116;

Matches 322; Conservative 82; Mismatches 104; Indels 28; Gaps 6;

17 FTRFMSTNKHDKIGLLYLVAAGVGTISLVFVVRLEMDPGVQYMC 76  
 4 FAQRWLFSTNKHDKIGLLYLVAAGVGTISLVFVVRLEMDPGVQYMC 48  
 77 QCTANGHLLNNVYTHGILMFVFGIPALFGSGFVNLPLQIGADMAF 136

Db 49 QVLGNHOLYNVITTAHAFLEFWMFPAALLGGFCNMFVPLMIGAPDMAEPRLNNISFWL 108  
 QY 137 FIAGTANGVASLAFARGDGOGLSGVGVNLYPLST--REAGYMDLAFVAHLSGASSIM 194  
 Db 109 LPPSLLLLLLSALV-----EIGAGTGWTVYPPPLSSIASGASVDLAFSLHISGASSIL 163  
 QY 195 GAINMTTFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVLAGATMMLTDNRNGTTF 254  
 Db 164 GAINMTTFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVLAGATMMLTDNRNGTTF 223  
 QY 255 FNPAGGDPPLLYOHLEFMEFGEHPEVYIILDPFGILSHVSTESKRPVGYLPMVYAMAI 314  
 Db 224 FDPAGGDPPLLYOHLEFMEFGEHPEVYIILDPFGILSHVSTESKRPVGYLPMVYAMAI 283  
 QY 315 GVLFGEVVAHMYTVGMSLTQOSEYMLATVAVPTGKIFSWIATMNGSVSEKSPMLM 374  
 Db 284 GILGFTVAHMYTVGMSLTQOSEYMLATVAVPTGKIFSWIATMNGSVSEKSPMLM 343  
 QY 375 AFGEFMEFLTVGVTGIVLAQAGLDRAVDHYVVAHFHYVMSLGAIFAGIYFMPKF 434  
 Db 344 AIGFLEFLTVGVTGIVLAQAGLDRAVDHYVVAHFHYVMSLGAIFAGIYFMPKF 403  
 QY 435 SGRAPEMAAKLHEWTFEFGANVTFEPQHFLGROGMRRYIDYEPALMKNKVSYGAF 494  
 Db 404 CGLOISETLGOIHMFVGVNLTFFPMHFLGISGMRRIDYDPAFAGMAVCSGSYL 463  
 QY 495 AFASFLEFVIVTVYVTVAGRRETRPNP--GEFADTLEWTLSPPAHFFELP 546  
 Db 464 SVLGALEFFVYV-VDTLTSGEERCPNRPWETVPAGISA-TLEMLTSPRAYHFFEEVP 517

RESULT 7

cytochrome-c oxidase (EC 1.9.3.1) chain I - liverwort (Marchantia polymorpha) mitochondrion  
 C:Species: mitochondrion Marchantia polymorpha  
 C>Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 23-Jul-1999  
 C:Accession: S25956  
 R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kar  
 J. Mol. Biol. 223, 1-7, 1992  
 A:Title: Gene organization deduced from the complete sequence of liverwort Marchantia po  
 A:Reference number: S25941; MUID:92114051  
 A:Accession: S25956  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-522 <D>A>  
 A:Cross-references: EMBL:M68929; NID:9786182; PIDD:AC09451.1; PIDD:9786237  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992  
 C:Genetics:  
 A:Gene: cox1  
 A:Genome: mitochondrion  
 A:Introns: 15/2; 60/1; 125/3; 132/2; 171/1; 208/3; 243/3; 372/3; 435/3  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as  
 transmembrane protein  
 F:13-460/Domain: cytochrome-c oxidase chain I homology <COI>  
 F:65,381/Binding site: heme a iron (His) (axial ligands) #status predicted  
 F:244,293,294/Binding site: copper (His) #status predicted  
 F:244-248/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:248/Binding site: oxygen (Tyr) #status predicted  
 F:371/Binding site: magnesium (His) (shared with chain II) #status predicted  
 F:379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 58.5%; Score 1756.5; DB 2; Length 522;  
 Best Local Similarity 58.9%; Pred. No. 9.2e-115;  
 Matches 315; Conservative 92; Mismatches 101; Indels 27; Gaps 6;  
 QY 17 FETRFMSTNKKDGLGLYLVAAGVGFISVLFVYMRLEMDPEVOYMCLEGARLIADASQC 76  
 Db 4 FAQRALFSTNKKDGLGLYLVAAGVGFISVLFVYMRLEMDPEVOYMCLEGARLIADASQC 48  
 QY 77 QCTANGHLNWNVTVYHGLIMFVGVIPALFGGNGNLTPLQCAPMAPPRMNNLSFWL 136  
 Db 1 I:||||| I:||||| I:||||| I:||||| I:||||| I:||||| I:||||| I:|||||

Db 49 QILGNHOLYNVITTAHAFLEFWMFPAALLGGFCNMFVPLMIGAPDMAEPRLNNISFWL 108  
 QY 137 FIAGTANGVASLAFARGDGOGLSGVGVNLYPLST--TREAGYMDLAFVAHLSGASSIM 194  
 Db 109 LPPSLLLLLLSALV-----EVGCGSGMTVYPPPLSSIASGASVDLAFSLHISGASSIL 163  
 QY 195 GAINMTTFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVLAGATMMLTDNRNGTTF 254  
 Db 164 GAINMTTFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVLAGATMMLTDNRNGTTF 223  
 QY 255 FNPAGGDPPLLYOHLEFMEFGEHPEVYIILDPFGILSHVSTESKRPVGYLPMVYAMAI 314  
 Db 224 FDPAGGDPPLLYOHLEFMEFGEHPEVYIILDPFGILSHVSTESKRPVGYLPMVYAMAI 283  
 QY 315 GVLFGEVVAHMYTVGMSLTQOSEYMLATVAVPTGKIFSWIATMNGSVSEKSPMLM 374  
 Db 284 GILGFTVAHMYTVGMSLTQOSEYMLATVAVPTGKIFSWIATMNGSVSEKSPMLM 343  
 QY 375 AFGEFMEFLTVGVTGIVLAQAGLDRAVDHYVVAHFHYVMSLGAIFAGIYFMPKF 434  
 Db 344 AIGFLEFLTVGVTGIVLAQAGLDRAVDHYVVAHFHYVMSLGAIFAGIYFMPKF 403  
 QY 435 SGRAPEMAAKLHEWTFEFGANVTFEPQHFLGROGMRRYIDYEPALMKNKVSYGAF 494  
 Db 404 TGLOYPETTLGOIHMFVGVNLTFFPMHFLGISGMRRIDYDPAFAGMAVCSGSYL 463  
 QY 495 AFASFLEFVIVTVYVTVAGRRETRPNP--GEFADTLEWTLSPPAHFFELP 546  
 Db 464 SVVGIFFEFVYV-VDTLTSGEERCPNRPWETVPAGISA-TLEMLTSPRAYHFFEEVP 516

RESULT 8

cytochrome-c oxidase (EC 1.9.3.1) chain I - sugar beet mitochondrion  
 C:Species: mitochondrion Beta vulgaris var. altissima (sugar beet)  
 C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 23-Jul-1999  
 C:Accession: S14138  
 R:Send, M.; Harada, T.; Mikami, T.; Sugijura, M.; Kinoshita, T.  
 Curr. Genet. 19, 175-181, 1991  
 A:Title: Genomic organization and sequence analysis of the cytochrome oxidase subunit  
 A:Reference number: S14138; MUID:91330331  
 A:Accession: S14138  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-524 <SEN>  
 A:Cross-references: GB:X57693; GB:S47702; NID:911258; PIDD:CAA0874.1; PIDD:911259  
 C:Genetics:  
 A:Gene: cox1  
 A:Genome: mitochondrion  
 A:Introns: 15/2; 60/1; 125/3; 132/2; 171/1; 208/3; 243/3; 372/3; 435/3  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane  
 transmembrane protein  
 F:12-459/Domain: cytochrome-c oxidase chain I homology <COI>  
 F:64,380/Binding site: heme a iron (His) (axial ligands) #status predicted  
 F:243,292,293/Binding site: copper (His) #status predicted  
 F:243-247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:247/Binding site: oxygen (Tyr) #status predicted  
 F:370/Binding site: magnesium (His) (shared with chain II) #status predicted  
 F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 59.3%; Score 1748.5; DB 2; Length 524;  
 Best Local Similarity 59.2%; Pred. No. 3.3e-114;  
 Matches 315; Conservative 86; Mismatches 106; Indels 25; Gaps 5;  
 QY 20 RFWFSTNKKDGLGLYLVAAGVGFISVLFVYMRLEMDPEVOYMCLEGARLIADASQC 79  
 Db 6 RFWFSTNKKDGLGLYLVAAGVGFISVLFVYMRLEMDPEVOYMCLEGARLIADASQC 50  
 QY 80 TANGHLNWNVTVYHGLIMFVGVIPALFGGNGNLTPLQCAPMAPPRMNNLSFWL 139  
 Db 51 GGNHOLYNVITTAHAFLEFWMFPAALLGGFCNMFVPLMIGAPDMAEPRLNNISFWL 110  
 QY 1 I:||||| I:||||| I:||||| I:||||| I:||||| I:||||| I:||||| I:|||||











C:Species: mitochondrion Zea mays (maize)  
 C:Date: 30-Jun-1992 #sequence.revision 30-Jun-1992 #text\_change 23-Jul-1999  
 C:Accession: A22840  
 R:Issac, P.G.; Jones, V.P.; Leaver, C.J.  
 EMBL J. 4, 1617-1623, 1985  
 A:Title: The maize cytochrome c oxidase subunit I gene: sequence, expression and rearing  
 A:Reference number: A22840  
 A:Accession: A22840  
 A:Molecule type: DNA  
 A:Residues: 1-528 <RSS>  
 A:Cross-references: EMBL:X02660; NID:g12888; PIDN:CAA26496.1; PID:g12889  
 A:Note: the authors translated the codon CGG for residue 405 as Trp, assuming a special  
 C:Genetics:  
 A:Gene: COXI  
 A:Genome: mitochondrion  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as  
 transmembrane protein  
 F:12-453/Domain: cytochrome-c oxidase chain I homology <COI>  
 F:64,380/Binding site: heme a iron (His) (axial ligands) #status predicted  
 F:243,292,293/Binding site: copper (His) #status predicted  
 F:243-247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:247/Binding site: oxygen (Tyr) #status predicted  
 F:370/Binding site: magnesium (His) (shared with chain II) #status predicted  
 F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 57.7%; Score 1730.5; DB 1; Length 528;  
 Best Local Similarity 58.5%; Pred. No. 5.9e-113;  
 Matches 311; Conservative 88; Mismatches 108; Indels 25; Gaps 5;

20 RWMSTNHKDIGLLYLVAGVGFISVLFYVYRLEMLDPGVQYMCLEGARLIADASQTC 79  
 Db RWFSTNHKDIGLLYLVAGVGFISVLFYVYRLEMLDPGVQYMCLEGARLIADASQTC 79  
 6 RWFSTNHKDIGLLYLVAGVGFISVLFYVYRLEMLDPGVQYMCLEGARLIADASQTC 79  
 QY 80 TANGHLMNVVYTYGILMFPVGFIPALFGFGNLYMLPLOGAPMAFPKNNISFWLFIA 139  
 Db 51 GGNHQLYNVLTAAHAFIMFPMVPMAMIGGFMFVILIGAPDMAPPRNNISFWLPP 110  
 QY 140 GTAGVASLFAFGGDLGSGVGVWLYPLPS--TREAGYSMDLAIFVHLGASSIMGAI 197  
 Db 111 SLLLSLSALV-----EVGSGTGTVYVPLPSGITSHSGAVDLAIFSLHLSGVSSILGSI 165  
 QY 198 NMITFLNMRAPGKTLHKVPLFSVIFITAMVLLALPVLAGATIMLTDNRNFTFEPN 257  
 Db 166 NFITTFNNMRGPGMTMHRPLFVWSVLTAFLLSLPVLGATIMLTDNRNFTFEPD 225  
 QY 258 AGGDDPLLYOHILMFEGHPEVYIIILPGFGLSHVSTFESKRPVGVLPVYVMAVIGVL 317  
 Db 226 AGGDDPLLYOHILMFEGHPEVYIIILPGFGLSHVSTFESKRPVGVLPVYVMAVIGVL 285  
 QY 318 GFVVAHHMYTVGMSLTQOQSYFMLATVAVPTGIKIFSMIATMGGSVFESKRPMLAFG 377  
 Db 286 GFVVAHHMYTVGMDVDTAYTAATMIIVPTGIKIFSMIATMGGSIQYKTPMLFAVG 345  
 QY 378 FMFLFVGGVTVGLVLAQADIRAYHDYVVAHFVYVMSLGAIFAFAGIYFMPKFSGR 437  
 Db 346 FIFLFTIGLGTGLVLANSLDIALHDYVVAHFVYVMSLGAIFAFAGIYFMPKFSGR 405  
 QY 438 AFPEMAAKLHFWTFEGICANTYFFPOHFLGRQGMRRYIDPEAFALMKNVSSYGAFLAFA 497  
 Db 406 TYPELGOIHFWITFEQVNLTFEPMHFLGLSGMPRRIDPDYAGVMAVLSFGYSISV 465  
 QY 498 SLFELFVIVYTVLVAGR--RETRPNPWG--EFADTLEWLPSPRPAHNTETPL 546  
 Db 466 GIRREFVVAITSSSGKNKCAESPMAVEQNPTTLEWLVQSPPAHNTETPL 517

RESULT 15  
 S16256  
 cytochrome-c oxidase (BC 1.9.3.1) chain I - wheat mitochondrion  
 C:Species: mitochondrion Triticum aestivum (common wheat)  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 23-Jul-1999

C:Accession: S16256; A26804; S33127  
 R:Rathburn, H.B.; Hedgcock, C.  
 Plant Mol. Biol. 16, 909-912, 1991  
 A:Title: A chimeric open reading frame in the 5' flanking region of cox1 mitochondria  
 A:Reference number: S16255; M0ID:91316228  
 A:Accession: S16256  
 A:Molecule type: DNA  
 A:Residues: 1-524 <RAT>  
 A:Cross-references: EMBL:X56186; NID:g13694; PIDN:CAA39651.1; PID:g13696  
 A:Note: the authors translated the codon CGG for residue 405 as Trp, assuming a special  
 R:Bonen, L.; Boer, P.H.; Moirans, J.E.; Gray, M.W.  
 Nucleic Acids Res. 15, 6734, 1987  
 A:Title: Nucleotide sequence of the wheat mitochondrial gene for subunit I of cytochr  
 A:Reference number: A26804; M0ID:87316932  
 A:Accession: A26804  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-404, 'W', 406-524 <BON>  
 A:Cross-references: GB:Y00417; NID:g13688  
 C:Genetics:  
 A:Gene: coxi  
 A:Genome: mitochondrion  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane  
 transmembrane protein  
 F:12-459/Domain: cytochrome-c oxidase chain I homology <COI>  
 F:64,380/Binding site: heme a iron (His) (axial ligands) #status predicted  
 F:243,292,293/Binding site: copper (His) #status predicted  
 F:243-247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F:247/Binding site: oxygen (Tyr) #status predicted  
 F:370/Binding site: magnesium (His) (shared with chain II) #status predicted  
 F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 57.6%; Score 1729.5; DB 2; Length 524;  
 Best Local Similarity 58.5%; Pred. No. 6.9e-113;  
 Matches 311; Conservative 89; Mismatches 107; Indels 25; Gaps 5;

20 RWMSTNHKDIGLLYLVAGVGFISVLFYVYRLEMLDPGVQYMCLEGARLIADASQTC 79  
 Db 6 RWFSTNHKDIGLLYLVAGVGFISVLFYVYRLEMLDPGVQYMCLEGARLIADASQTC 79  
 QY 80 TANGHLMNVVYTYGILMFPVGFIPALFGFGNLYMLPLOGAPMAFPKNNISFWLFIA 139  
 Db 51 GGNHQLYNVLTAAHAFIMFPMVPMAMIGGFMFVILIGAPDMAPPRNNISFWLPP 110  
 QY 140 GTAGVASLFAFGGDLGSGVGVWLYPLPS--TREAGYSMDLAIFVHLGASSIMGAI 197  
 Db 111 SLLLSLSALV-----EVGSGTGTVYVPLPSGITSHSGAVDLAIFSLHLSGVSSILGSI 165  
 QY 198 NMITFLNMRAPGKTLHKVPLFSVIFITAMVLLALPVLAGATIMLTDNRNFTFEPN 257  
 Db 166 NFITTFNNMRGPGMTMHRPLFVWSVLTAFLLSLPVLGATIMLTDNRNFTFEPD 225  
 QY 318 GFVVAHHMYTVGMSLTQOQSYFMLATVAVPTGIKIFSMIATMGGSVFESKRPMLAFG 377  
 Db 286 GFVVAHHMYTVGMDVDTAYTAATMIIVPTGIKIFSMIATMGGSIQYKTPMLFAVG 345  
 QY 378 FMFLFVGGVTVGLVLAQADIRAYHDYVVAHFVYVMSLGAIFAFAGIYFMPKFSGR 437  
 Db 346 FIFLFTIGLGTGLVLANSLDIALHDYVVAHFVYVMSLGAIFAFAGIYFMPKFSGR 405  
 QY 438 AFPEMAAKLHFWTFEGICANTYFFPOHFLGRQGMRRYIDPEAFALMKNVSSYGAFLAFA 497  
 Db 406 TYPELGOIHFWITFEQVNLTFEPMHFLGLSGMPRRIDPDYAGVMAVLSFGYSISV 465  
 QY 498 SLFELFVIVYTVLVAGR--RETRPNPWG--EFADTLEWLPSPRPAHNTETPL 546  
 Db 466 GIRREFVVAITSSSGKNKCAESPMAVEQNPTTLEWLVQSPPAHNTETPL 517

Thu Dec 27 08:21:34 2001

us-09-712-768-2.rpr

Page 10

Search completed: December 22, 2001, 11:02:25  
Job time: 424 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:02:31 ; Search time 15.26 seconds

(Without alignments)  
1338.290 Million cell updates/sec

Title: US-09-712-768-2  
Perfect score: 3000  
Sequence: 1 MADDAIHGHDKHKGQFTR.....PAHTEPLPKRSDMKHPHS 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2538.5	84.6	558	1	CX1B_PARDE
2	2453	81.8	565	1	COX1_RHOSH
3	2347	78.2	554	1	CX1A_PARDE
4	1873.5	62.5	538	1	COX1_RHLE
5	1829	58.5	541	1	COX1_BRAJA
6	1756.5	58.5	522	1	COX1_MARPO
7	1748.5	58.3	524	1	COX1_BENVU
8	1744	58.1	534	1	COX1_RICPR
9	1738.5	58.0	527	1	COX1_ARATH
10	1730.5	57.7	524	1	COX1_ORYZA
11	1730.5	57.7	528	1	COX1_ORYZA
12	1730.5	57.6	530	1	COX1_SORBI
13	1729.5	57.6	524	1	COX1_WHEAT
14	1713.5	57.1	514	1	COX1_PROMT
15	1677.5	55.9	532	1	COX1_CHOCR
16	1677	55.9	536	1	COX1_ALIMA
17	1651.5	55.0	527	1	COX1_PEA
18	1643.5	54.8	527	1	COX1_SOYBN
19	1639	54.6	530	1	COX1_METSE
20	1609	53.6	557	1	COX1_NEUCR
21	1608.5	53.6	527	1	COX1_OENBE
22	1605.5	53.5	528	1	COX1_EMENT
23	1600	53.3	528	1	COX1_TRIRU
24	1594	53.1	526	1	COX1_CVACA
25	1594	53.1	541	1	COX1_PODAN
26	1593	53.1	492	1	COX1_PHYME
27	1588.5	53.0	517	1	COX1_PARLI
28	1564.5	52.1	517	1	COX1_STRPU
29	1562.5	52.1	517	1	COX1_PISOC
30	1561.5	52.0	873	1	COX1_ACACA
31	1559	52.0	517	1	COX1_MYXGL
32	1558.5	52.0	517	1	COX1_ASTPE
33	1554.5	51.8	512	1	COX1_YEAST

34	1550	51.7	513	1	COX1_RABIT	079429 oryctolagus
35	1547	51.6	516	1	COX1_CARAU	078681 catassius a
36	1546	51.5	516	1	COX1_GADMO	Q36775 gadus morhu
37	1545.5	51.5	534	1	COX1_SACDO	P98001 saccharomyc
38	1545	51.5	516	1	COX1_STRCA	021399 struthio ca
39	1544	51.5	514	1	COX1_CERSI	003198 ceratotheri
40	1544	51.5	514	1	COX1_SHEEP	078749 ovis aries
41	1543	51.4	513	1	COX1_DIDMA	P41310 dielphis m
42	1543	51.4	514	1	COX1_BOVIN	P00396 bos taurus
43	1543	51.4	516	1	COX1_CROLA	P34188 crotosoma
44	1542	51.4	514	1	COX1_FELCA	P48888 felis silve
45	1542	51.4	514	1	COX1_RAT	P05503 rattus norv

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	558 AA.
1	CX1B_PARDE			
AC	P98002;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CYTOCHROME C OXIDASE POLYPEPTIDE I-BETA (EC 1.9.3.1) (CYTOCHROME AA3 SUBUNIT I-BETA).			
GN	CTADII.			
OS	Paracoccus denitrificans.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;			
OC	Paracoccus.			
OX	NCBI_TaxID=266;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-PD 1222;			
RX	MEDLINE=90184495; PubMed=2155830;			
RA	Raitio M., Pispas J.M., Melse T., Saraste M.;			
RT	"Are there isoenzymes of cytochrome c oxidase in Paracoccus denitrificans?";			
RT	FEBS Lett. 261:431-435(1990).			
RL	[2]			
RN	X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).			
RP	MEDLINE=95379947; PubMed=7651515;			
RX	Iwata S., Ostermeier C., Ludwig B., Michel H.;			
RA	"Structure at 2.8-A resolution of cytochrome c oxidase from Paracoccus denitrificans.";			
RT	Nature 376:660-669(1995).			
RL	[3]			
RN	X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).			
RP	MEDLINE=98021406; PubMed=9380672;			
RX	Ostermeier C., Harrenga A., Ermler U., Michel H.;			
RA	"Structure at 2.7-A resolution of the Paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody FV fragment.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 94:10547-10553(1997).			
RL	[4]			
RN	COVALENT BOND.			
RP	MEDLINE=99268331; PubMed=10338009;			
RX	Buse G., Soullimane T., Dewor M., Meyer H.E., Blueggel M.;			
RA	"Evidence for a copper-coordinated histidine-tyrosine cross-link in the active site of cytochrome oxidase.";			
RT	Protein Sci. 8:985-990(1999).			
RL	[5]			
RN	REVIEW.			
RP	MEDLINE=96391941; PubMed=8794157;			
RX	Ostermeier C., Iwata S., Michel H.;			
RA	"Cytochrome c oxidase.";			
RT	Curr. Opin. Struct. Biol. 6:460-466(1996).			
RL	[6]			
CC	-I- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND CU(B). THIS CYTOCHROME C OXIDASE SHOWS PROTON PUMP ACTIVITY ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER.			
CC	-I- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +			

CC 4 FERRICYTOCHROME C.  
 CC -1- COFACTOR: TWO HEME GROUPS AND COPPER B.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, INNER MEMBRANE.  
 CC -1- PTM: HIS-276 AND TYR-280 RESIDUES ARE INVOLVED IN THE FORMATION OF  
 CC A COPPER-COORDINATED COVALENT CROSSLINK AT THE ACTIVE SITE OF THE  
 CC CATALYTIC SUBUNIT 1.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Y07533; CAA68821.1; -  
 CC PIR: S08270; S08270.  
 CC PDB: 1AR1; 1I-FEB-98.  
 CC InterPro: IPR000883; COX1.  
 CC Pfam: PF00115; COX1; 1.  
 CC PRINTS: PR01165; CYCOXIDASE1.  
 CC PROSITE: PS00077; COX1; 1.  
 CC Oxioreductase; Heme; Copper; Transmembrane; Inner membrane;  
 CC Respiratory chain; Hydrogen ion transport; 3D-structure.  
 CC  
 CC FT DOMAIN 1 28  
 CC FT TRANSMEM 29 59  
 CC FT TRANSMEM 60 82  
 CC FT TRANSMEM 83 120  
 CC FT DOMAIN 121 126  
 CC FT TRANSMEM 127 151  
 CC FT DOMAIN 152 176  
 CC FT TRANSMEM 177 206  
 CC FT DOMAIN 207 217  
 CC FT TRANSMEM 218 251  
 CC FT DOMAIN 252 262  
 CC FT TRANSMEM 263 299  
 CC FT DOMAIN 300 303  
 CC FT TRANSMEM 304 319  
 CC FT DOMAIN 320 332  
 CC FT TRANSMEM 333 364  
 CC FT DOMAIN 365 369  
 CC FT TRANSMEM 370 395  
 CC FT DOMAIN 405 404  
 CC FT TRANSMEM 405 437  
 CC FT DOMAIN 438 440  
 CC FT TRANSMEM 441 469  
 CC FT DOMAIN 470 478  
 CC FT TRANSMEM 479 514  
 CC FT DOMAIN 515 558  
 CC FT METAL 94 94  
 CC FT METAL 276 276  
 CC FT METAL 280 280  
 CC FT METAL 325 325  
 CC FT METAL 326 326  
 CC FT METAL 411 411  
 CC FT METAL 413 413  
 CC FT DISULFID 66 80  
 CC FT SEQUENCE 558 AA; 62439 MW; A8402453C0C0339E CRC64;

Query Match 84.6%; Score 2538.5; DB 1; Length 558;  
 Best local Similarity 81.0%; Pred. No. 1.7e-140;  
 Matches 452; Conservative 51; Mismatches 54; Indels 1; Gaps 1;

QY 1 MADAAIGH-DHHEKOGFTFRMFSNKHDIIGLLIYVAGVGVISVLTFTVRLLELMP 59  
 DB 1 MADAAYGHGDHDTGRGFTFRMFSNKHDIIGLLIYFTAGVGLISVCFVYRMELQHP 60  
 QY 60 GVOYMCLEGARLLADASQCTANGHLMNWNVYTHGLIMEFVGIPLAFGFGNYLPLQI 119  
 DB 61 GVOYMCLEGARLLADASQCTANGHLMNWNVYTHGLIMEFVGIPLAFGFGNYLPLQI 120

QY 120 GAPDAFPRMNNLSFWLEIAGTAMGVASLAFAPGGDGLSGVGVNLVPLSTREAGYSMD 179  
 DB 121 GAPDAFPRMNNLSFWLEIAGTAMGVASLAFAPGGDGLSGVGVNLVPLSTREAGYSMD 180  
 QY 180 LAIFAVHLSGASSIMGAINMTTFLNMRAPDWTLHKVPLFSWIFITAMLLALPVLAG 239  
 DB 181 LAIFAVHLSGASSIMGAINMTTFLNMRAPDWTLHKVPLFSWIFITAMLLALPVLAG 240  
 QY 240 AITMLTLTRNCTFFFNAGGDDPLVONILMERGHEVYIIIPGCIISHVSTSSK 299  
 DB 241 AITMLTLTRNCTFFFNAGGDDPLVONILMERGHEVYIIIPGCIISHVSTSSK 300  
 QY 300 PVFGYLVPMVAVIAGVGFVMAHMTVMSLTQOSYFMLATVAVPTGKIFSMIA 359  
 DB 301 PIFGLPVLVMAIIGLGFVMAHMTVMSLTQOSYFMLATVAVPTGKIFSMIA 360  
 QY 360 TWMGSGVEFKSPMLWAGFMFLTVGVGTIVLAQAGLDRAVHDYTVVAHFVMSLGA 419  
 DB 361 TWMGSGVEFKSPMLWAGFMFLTVGVGTIVLAQAGLDRAVHDYTVVAHFVMSLGA 420  
 QY 420 TFAIFAGIYFYMFKPSGRAPPEMAKLFHFWTFEIGAVTFFPOHFLRGCPRRYIDYE 479  
 DB 421 VEGTLAGVYVWIGKSGQYPEMAGQLHEMMFTGSLNLFEPQHLRGQPPRYIDYE 480  
 QY 480 AFALMNKVSYSYGAFLAFASFLFVIVYTLVAGRRETRPMPGEPADTLEWTLSPSPA 539  
 DB 481 EFAYVNNNSISGAVISFASFLFVIVYTLVAGRRETRPMPGEPADTLEWTLSPSPA 540  
 QY 540 HTEFLPKRSMDKHPSH 557  
 DB 541 HTEFLPKRSMDKHPSH 558

RESULT 2  
 COX1\_RHOSH STANDARD; PRT; 565 AA.  
 AC P3517;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME AA3  
 SUBUNIT 1).  
 GN CTAD.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 OS Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;  
 OC Rhodospirillum.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GA;  
 RX MEDLINE=92204019; PubMed=1313140;  
 RA Shapleigh J.P., Genis R.B.;  
 RT "Cloning, sequencing and deletion from the chromosome of the gene  
 RT encoding subunit I of the aa3-type cytochrome c oxidase of  
 RT Rhodospirillum rubrum".  
 RL Mol. Microbiol. 6:635-642(1992).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B. THIS CYTOCHROME C OXIDASE SHOWS PROTON PUMP ACTIVITY  
 CC ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
 CC 4 FERRICYTOCHROME C.  
 CC -1- COFACTOR: TWO HEME GROUPS AND COPPER B.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 12  
 CC POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X62645; CAA44514.1; -  
DR PIR: S20534; S20534.  
DR HSSP: P98002; IARI.  
DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS: PR01165; COX1DASE1.  
DR PROSITE: PS00077; COX1; 1.  
KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;  
KW Hydrogen ion transport.  
FT METAL 102 102 IRON (HEME A) (PROBABLE).  
FT METAL 284 284 COPPER B (PROBABLE).  
FT METAL 288 288 COPPER B (PROBABLE).  
FT METAL 333 333 COPPER B (PROBABLE).  
FT METAL 334 334 COPPER B (PROBABLE).  
FT METAL 419 419 IRON (HEME A3) (PROBABLE).  
FT METAL 421 421 IRON (HEME A) (PROBABLE).  
SQ SEQUENCE 565 AA; 62685 MW; A74738C138DA04E1 CRC64;

Query Match 81.8%; Score 2453; DB 1; Length 565;  
Best Local Similarity 78.7%; Pred. No. 1.5e-135; Mismatches 59; Indels 12; Gaps 3;  
Matches 446; Conservative 50;

QY 1 MADAAIHGHHKOGFTFMNSTNKHDIQLYLVAAGVFTSVLTFTYVRLMDPG 60  
DB 1 MADAAIHGHE-HDRRGFTFMNSTNKHDIQLYLVAAGVFTSVLTFTYVRLMDPG 59  
QY 61 VOYKLEBAR-----LIADSOCTANGHLNMYVTHGLIMFEYGTALGCGF 110  
DB 60 VOYKLEBESGLVKGFQFOSLPSAVENCSPNGHLMVMTITGHILMFEVYVIALGCGF 119  
QY 111 GNYLMPDLIGADMAFPFRNNLSFWLFTAGTAMGVASLFARFGGOLSGGVWLYPPLS 170  
DB 120 GNYLMPDLIGADMAFPFRNNLSFWLFTAGTAMGVASLFARFGGOLSGGVWLYPPLS 179  
QY 171 TREAGYSMDLAIFAVHLSGASSIGALIMTTFINMRAPGWTLHKVPLFSWIFITAMLI 230  
DB 180 TSESGYSTDLAIFAVHLSGASSIGALIMTTFINMRAPGWTLHKVPLFSWIFITAMLI 239  
QY 231 LLALPYLAGATITMLLDNENFTFNPAGGDDPLIXOIIIMFEGHREYVYIIILGFGIIS 290  
DB 240 LLALPYLAGATITMLLDNENFTFNPAGGDDPLIXOIIIMFEGHREYVYIIILGFGIIS 299  
QY 291 HVSTFSKRPVEGYLPRVYAVVAIGVLGFVYVWVHNMVTVGMSLQOQSFMLATVIAVPT 350  
DB 300 HVSTFSKRPVEGYLPRVYAVVAIGVLGFVYVWVHNMVTVGMSLQOQSFMLATVIAVPT 359  
QY 351 GIKIFSWATWYMGSSVEKSPMLAFGFMELFTVGVGTGYLAQAGIDRAYHDTYVVAH 410  
DB 360 GIKIFSWATWYMGSSVEKSPMLAFGFMELFTVGVGTGYLAQAGIDRAYHDTYVVAH 419  
QY 411 FHYVMSLGATIFAGIYFTVMPKFSGRAPRMAKLFHMTFFIGANTYFRPHNLGQGM 470  
DB 420 FHYVMSLGATIFAGIYFTVMPKFSGRAPRMAKLFHMTFFIGANTYFRPHNLGQGM 479  
QY 471 PRVYIDPEAFALMNKVSYSYGAFALFASLEFFIYFYVTVAGRRETRPNMGSEFADTLE 530  
DB 480 PRVYIDPEAFALMNKVSYSYGAFALFASLEFFIYFYVTVAGRRETRPNMGSEFADTLE 538  
QY 531 WTLSPPPAHFFETLPKRSDMKHPHSH 557  
DB 539 WTLSPPPAHFFETLPKRSDMKHPHSH 565

RESULT 3

CX1A\_PARDE  
ID CX1A\_PARDE STANDARD; PRT; 554 AA.  
AC P08305;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I-ALPHA (EC 1.9.3.1) (CYTOCHROME AA3  
DE SUBUNIT I-ALPHA).  
GN CYADI OR COI.  
OS Paracoccus denitrificans.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Paracoccus.  
OX NCBI\_TaxID=266;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SI657;  
RA Raitio M., Jalli T., Saraste M.;  
RT "Isolation and analysis of the genes for cytochrome c oxidase in  
RT Paracoccus denitrificans."  
RL EMBO J. 6:2825-2833(1987).  
CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME  
CC COMPLEX. ELECTROMS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA  
CC CU(B). THIS CYTOCHROME C OXIDASE SHOWS PROTON PUMP ACTIVITY  
CC ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER.  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +  
CC 4 FERROCYTOCHROME C.  
CC -1- COFACTOR: TWO HEME GROUPS AND COPPER B.  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X05829; CAA29274.1; -  
DR PIR: S03809; S03809.  
DR HSSP: P98002; IARI.  
DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS: PR01165; CYCOXIDASE1.  
DR PROSITE: PS00077; COX1; 1.  
KW Oxidoreductase; Heme; Copper; Transmembrane; Inner membrane;  
KW Respiratory chain; Hydrogen ion transport.  
FT METAL 91 91 IRON (HEME A) (BY SIMILARITY).  
FT METAL 273 273 COPPER B (BY SIMILARITY).  
FT METAL 277 277 COPPER B (BY SIMILARITY).  
FT METAL 322 322 COPPER B (BY SIMILARITY).  
FT METAL 323 323 COPPER B (BY SIMILARITY).  
FT METAL 406 406 IRON (HEME A3) (BY SIMILARITY).  
FT METAL 408 408 IRON (HEME A) (BY SIMILARITY).  
FT DISULFID 63 63 BY SIMILARITY.  
FT TRANSMEM 26 56 BY SIMILARITY.  
FT TRANSMEM 81 118 BY SIMILARITY.  
FT TRANSMEM 127 148 BY SIMILARITY.  
FT TRANSMEM 175 203 BY SIMILARITY.  
FT TRANSMEM 215 248 BY SIMILARITY.  
FT TRANSMEM 260 295 BY SIMILARITY.  
FT TRANSMEM 301 319 BY SIMILARITY.  
FT TRANSMEM 331 359 BY SIMILARITY.  
FT TRANSMEM 367 390 BY SIMILARITY.  
FT TRANSMEM 399 425 BY SIMILARITY.  
FT TRANSMEM 436 463 BY SIMILARITY.  
FT TRANSMEM 478 508 BY SIMILARITY.  
SQ SEQUENCE 554 AA; 62012 MW; 58AD591FBDCE794 CRC64;

Query Match

78.2%; Score 2347; DB 1; Length 554;

Best Local Similarity 76.1%; Pred. No. 2,1e-129;  
Matches 415; Conservative 59; Mismatches 69; Indels 2; Gaps 1;

QY 13 EKGGFTTRPMSTNHRDGLLYLVAGVGTISVLTVMRLELMDPPGVYMCLEBARLI 72  
DB 11 EKGGFTTRPMSTNHRDGLLYLVAGVGTISVLTVMRLELMDPPGVYMCLEBARLI 70  
QY 73 ADASQTCNANGLHNMVNYHGLIMEFVGIPALPGCFNLYLMDLQIGADMAFRPMNL 132  
DB 71 ADAACCTPNAHNMVNYHGLIMEFVGIPALPGCFNLYLMDLQIGADMAFRPMNL 130  
QY 133 SEWFLIAGTAMGASLAFAPGDDQLSGVGVWVLPPLSTREAGYSMDLAFVHLSGASS 192  
DB 131 SYVLVYCGVSLATSLSPGSDQPGAGVGVWVLPPLSTREAGYAMDALFVHVSAGTS 190  
QY 193 IMCAIMWITTFELMRAPGKTLHKVPLFSISITFATLILALPYLAGATMLTDNRNFT 252  
DB 191 ILCAIMWITTFELMRAPGKTLHKVPLFSISITFATLILALPYLAGATMLTDNRNFT 250  
QY 253 TFFNPAGGDPILYOHILMFEGHPEVYIILPGFIISHVSTFSKPKFYGLPMVYANV 312  
DB 251 QFFDPAGGDPVLYOHILMFEGHPEVYIILPGFIISHVSTFSKPKFYGLPMVYANV 310  
QY 313 AIGVLGFVVAHHMYTVGMSLTQOQSYFMLATMVIAVPTGIKIFSWIATMGGSVFEKSPM 372  
DB 311 AIFLGFIVVAHHMYTVGMSLTQOQSYFMLATMVIAVPTGIKIFSWIATMGGSVFEKSPM 370  
QY 373 LMAFGMPLFTVGVGVYLAQAGLDRAVHDYVYVAHHMYMSGATFALFAGIYFVMP 432  
DB 371 LMAFLA--FLFTVGVGVYLAQAGLDRAVHDYVYVAHHMYMSGATFALFAGIYFVMP 428  
QY 433 KFSGRAPPEMAKHLHWTFFIGANTVFPOHFLRGOMPRRYIDYDEAFALNKKVSSYA 492  
DB 429 KMSGRQYPRBAOGLHWMFISGLNLFPPHFLRGOMPRRYIDYDEAFALNKKVSSYA 488  
QY 493 FLAFAFLFTFIVYTVLAGRRTRPNGEFADTLENTLDSPPRAHFEPLPKRSDMD 552  
DB 489 YISFAFLFTFIVYTVLAGRRTRPNGEFADTLENTLDSPPRAHFEPLPKRSDMD 548  
QY 553 KHRSH 557  
DB 549 RAOAH 553

RESULT 4  
COX1\_RHILE STANDARD: PRT: 538 AA.  
ID COX1\_RHILE  
AC 008855;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME AA3 SUBUNIT 1).  
GN CTAD OR COXA.  
OS Rhizobium leguminosarum.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
NCBI\_TaxID=384;  
RN 111  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94161496; PubMed=8117073;  
RT Gabel C., Bittenger M.A., Maiter R.J.;  
RT "Cytochrome a3 gene regulation in members of the family  
Rhizobiaceae: comparison of copper and oxygen effects in  
Bradyrhizobium japonicum and Rhizobium tropici.";  
RT Appl. Environ. Microbiol. 60:141-148(1994).  
RL -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3  
AND COPPER B.

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
CC 4 FERROCYTOCHROME C.  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 12  
CC POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- DEVELOPMENTAL STAGE: FREE IN SOIL (NOT AS BACTERIUM).  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: X74341; CA52388.1; -;  
DR HSSP: P98002; IARL.  
DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS: PR01165; CYCOXIDASEI.  
DR PROSITE: PS00077; COX1; 1.  
KM Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.  
FT TRANSSEM 37 57  
FT TRANSSEM 84 104  
FT TRANSSEM 121 141  
FT TRANSSEM 170 190  
FT TRANSSEM 208 228  
FT TRANSSEM 259 279  
FT TRANSSEM 291 311  
FT TRANSSEM 329 349  
FT TRANSSEM 362 382  
FT TRANSSEM 401 421  
FT TRANSSEM 438 458  
FT TRANSSEM 476 496  
FT METAL 82 82  
FT METAL 265 265  
FT METAL 269 269  
FT METAL 314 314  
FT METAL 315 315  
FT METAL 400 400  
FT METAL 402 402  
SQ SEQUENCE 538 AA; 59133 MW; 740C8F6C902D9BF CRC64;  
Query Match 62.5%; Score 1873.5; DB 1; Length 538;  
Best Local Similarity 64.2%; Pred. No. 5.3e-102;  
Matches 348; Conservative 63; Mismatches 110; Indels 21; Gaps 4;  
QY 7 HGHDRHEKQGFTRPMSTNHRDGLLYLVAGVGTISVLTVMRLELMDPPGVYMCLE 66  
DB 15 HGHDRHAPR--TGMRRSTRNHRDGLLYLVAGVGTISVLTVMRLELMDPPGVYMCLE 69  
QY 67 EGARLIADASQTCNANGLHNMVNYHGLIMEFVGIPALPGCFNLYLMDLQIGADMAFR 126  
DB 70 -----HNTLHLYVVFVTSHGVIIMIFPMVPMAGGGMVFLPMLGADPMAR 115  
QY 127 PRMNLSPWFLIAGTAMGASLAFAPGDDQLSGVGVWVLPPLSTREAGYSMDLAFV 185  
DB 116 PRMNLSPWFLIAGTAMGASLAFAPGDDQLSGVGVWVLPPLSTREAGYSMDLAFV 175  
QY 186 HSGASSTIGAGINFTITLNNRARGMTLHKVPLFSISITFATLILALPYLAGATML 245  
DB 176 HSGASSTIGAGINFTITLNNRARGMTLHKVPLFSISITFATLILALPYLAGATML 235  
QY 246 TDRNFGTTFEPNAGGDPVLYOHILMFEGHPEVYIILPGFIISHVSTFSKPKFYGL 305  
DB 236 TDRNFGTTFEPNAGGDPVLYOHILMFEGHPEVYIILPGFIISHVSTFSKPKFYGL 295  
QY 306 PMVYAMVAIGVLGFVVAHHMYTVGMSLTQOQSYFMLATMVIAVPTGIKIFSWIATMGG 365  
DB 296 GNVYAMVAIGVLGFVVAHHMYTVGMSLTQOQSYFMLATMVIAVPTGIKIFSWIATMGG 355



```

RESULT 6
COX1_MARPO STANDARD: PRT: 522 AA.
AC P26856;
DB 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Marchantia polymorpha (Liverwort).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiales; Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kaneage T., Ogura Y., Kohchi T., Ohyama K.;
RT "Gene organization deduced from the complete sequence of liverwort
RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
RT mitochondrial genome."
RL J. Mol. Biol. 223:1-7(1992).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL: M68929; AAC09451.1; -
CC PIR: S25956; S25956.
CC DR HSP: P00396; IOCC.
CC DR Wende: 2055; MARPO;cox1;1.
CC DR InterPro: IPR000883; COX1.
CC DR Pfam: PF00115; COX1; 1.
CC DR PRINTS: PR01165; CYCOXIDASE1.
CC DR PROSITE: PS00077; COX1; 1.
CC KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
CC Respiratory chain; Inner membrane.
CC FT METAL 65 65 IRON (HEME A) (PROBABLE).
CC FT METAL 244 244 COPPER B (PROBABLE).
CC FT METAL 248 248 COPPER B (PROBABLE).
CC FT METAL 293 293 COPPER B (PROBABLE).
CC FT METAL 294 294 COPPER B (PROBABLE).
CC FT METAL 379 379 IRON (HEME A3) (PROBABLE).
CC FT METAL 381 381 IRON (HEME A) (PROBABLE).
CC SQ SEQUENCE 522 AA; 57551 MW; 4B88E16F03A11D01 CRC64;

Query Match 58.5%; Score 1756.5; DB 1; Length 522;
Best Local Similarity 58.9%; Pred. No. 3,1e-95;
Matches 315; Conservative 92; Mismatches 101; Indels 27; Gaps 6;

QY 17 FFRWMTNHRKDIGLLTYAACGVFISLFTVYMRLEMDPGVGYMCLGRLADAS 76
DB 4 FAORWLPSTNHRKDIGLLTYAACGVFISLFTVYMRLEMDPGVGYMCLGRLADAS 76

```

```

QY 77 QTCANGHLNWNVTYHGLMMPFVGPALFPGFGNGYLMPLQIGADPAPMARRNNLSFWL 136
DB 49 QILGNGHQLVLLTAHAFLIFFWVPMAMIGCGNMFVILIGSPMAPPRLNNISFWL 108
QY 137 FIAGTACVASFAPGGDGLSGGVNLPPLS--TREACTSMDLAIFAVHLSGASSIM 194
DB 109 LPPSLILLSSALV----EVGCGSGMTVYPPSLGISYSHSGGSVDLAIFSLHSGVSL 163
QY 195 GAINMTTFLMRAPGMTHKVPFLPFSMIFTAFLILALPVLAGATMTLDRNFGTTF 254
DB 164 GSINFTITTFMRAPGLTMHRLPLFVMSVLTATLLSLPVLGATMTLDRNFGTTF 223
QY 255 FNPAGGGDPLLYOHLMEFGHPEVYIILDPGIIISHVSTFSKRPVGYLPMYAVNAVI 314
DB 224 FDPAGGGDPLLYOHLMEFGHPEVYIILDPGIIISHVSTFSKRPVGYLPMYAVNAVI 283
QY 315 GVLGFVVAHHMTYVNGSLTQGSFMTLATWIANVPTGIKIFSMTATMGGSVKESFML 374
DB 284 GVLGFVVAHHMTYVNGSLTQGSFMTLATWIANVPTGIKIFSMTATMGGSIQYKPMPLF 343
QY 375 AEGFMEFTVGVTGIVLAQAGLDRAVDGYVVAHHFVWMSLGAIFAIFAGIFYMPKE 434
DB 344 AVGFTEFTVGGLTGIYLANSGVDIALHDITYVVAHHFVWMSLGAIFAIFAGIFYWIGKI 403
QY 435 SGRAPEMAAKLHFWTFEIGANTVTFPDHFLGRGMRRYIDPEADALNNKVSSTGAF 494
DB 404 TGLQPEFTLGGIHFWTFEIGANTVTFPDHFLGRGMRRYIDPEADALNNKVSSTGAF 463
QY 495 AFAS-LEFFIYIFVYTLVAGRRTRPNWG--EFADLLEWTLSPPPAHFEETLP 546
DB 464 SVGICFCEVVFVFL--TLTSNKCASPFAVEONSTLEWNVSPPAHFEETLP 516

RESULT 7
COX1_BETVU STANDARD: PRT: 524 AA.
AC P24794;
DB 01-MAR-1992 (Rel. 21, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Becta vulgaris (Sugar beet).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Caryophyllales; Magnoliophyta; eudicotyledons; core eudicots;
OC NCBI_TaxID=3555;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, TK81-0;
RX MEDLINE=9130331; PubMed=1651175;
RA Senda M., Harada T., Mikami T., Sugura M., Kinoshita T.;
RT "Genomic organization and sequence analysis of the cytochrome oxidase
RT subunit II gene from normal and male-sterile mitochondria in sugar
RT beet."
RL Curr. Genet. 19:175-181(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Harada T., Mikami T., Kinoshita T.;
RT "Nucleotide sequence of cytochrome c oxidase subunit I gene of sugar
RT beet mitochondria."
RL Proc. Sugar Beet Res. 29:15-21(1987).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

```



```

CC      -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC      INNER MEMBRANE. CONSTRAINTS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC      -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its use
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb-sdb.ch).
CC
DR      EMBL: X57693; CAA40874.1; -.
DR      EMBL: M57645; AAA87330.1; -.
DR      HSSP: P00396; IOCC.
DR      Mendel: 5481; BFTvu:cox1.1.
DR      InterPro: IPR000883; COX1.
DR      Pfam: PF00115; COX1. 1.
DR      PRINTS: PR01165; CYCOXIDASE1.
DR      PROSITE: PS00077; COX1. 1.
KW      Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW      Respiratory chain; Inner membrane.
FT      METAL                     64
FT      METAL                     64
FT      METAL                     243
FT      METAL                     243
FT      METAL                     247
FT      METAL                     292
FT      METAL                     292
FT      METAL                     293
FT      METAL                     293
FT      METAL                     378
FT      METAL                     378
FT      METAL                     380
FT      METAL                     380
FT      CONFLICT                   2
FT      CONFLICT                   9
FT      CONFLICT                   220
FT      CONFLICT                   226
FT      CONFLICT                   229
FT      CONFLICT                   229
FT      CONFLICT                   501
FT      CONFLICT                   503
SQ      SEQUENCE                   524 AA; 57580 MW; A862089E5C476EA CRC64;

Query Match          58.3%; Score 1748.5; DB 1; Length 524;
Best Local Similarity 59.2%; Pred. No. 9e-95;
Matches 315; Conservative 86; Mismatches 106; Indels 25; Gaps 5.

QY      20  RWFMSTNRHKDGLGLYLVAAGVGFISLFTYVYRLMDPCGYQVMCLGARLIADASQC 79
Dd      6  RWLESTNRKDKDGLGLYLFGTALGAGVAGCTGFSYLIRKELARPG-----DQIL 50
QY      80  TANGHLMNVVNTYTGILMFEVGIIPALEGEGENYLMPIQIGADMAFPRMNLSEMLFIA 139
Dd      51  GGNHQLYVNLITAHAFLEIMFEVGMAMIGFGENFVPLIGADMAFPRLNISFMLLP 110
QY      140  GTAMVVALFLPAGGGGGLGSGVGNVLPPLS--TEAGYSMDLAIFAVNLGASSIMGAI 197
Dd      111  SLLLLSSALV-----EVGSGTGMVTVPLSGITSHSGGAVDLAIFSLHSGVSSITGSI 165
QY      198  NMITTFELMRAPGMLLHNVLPFSISFIPTAMVILALPVLACATMLLTDNRNGTTFENP 257
Dd      166  NFITTFELMRGPGMTMHLRPLFVWSVLTAFLLLSLPLACATMLLTDNRNFNTTFEDP 225
QY      258  AGGCDPILYQHLMEFEGHREYVLIILPGEGILSHAVSTFSKKRVEGSLPMVYAMAIVGL 317
Dd      226  AGGCDPILYQHLMEFEGHREYVLIILPGEGILSHIVTFESGKRPVGYGLMVYAMISIGVL 285
QY      318  GFVVAHHMITYGMSLLQOOSTYMLATVNIAYVTGKIFSWLITMNGGSEYKSPMIAFG 377
Dd      286  GFVVAHHMIFVTVGLDVRAYFAFTAAITMIAVPTGKIFSWLITMNGGSIQYTKTPIMLFAVG 345
QY      378  FMFLFTVGGVGVIGVLAQGLDRAAYHDYVVAHFVYVMSLGAFIAFGIYVMPKFSGR 437
Dd      346  FIFLFTVGGVGVIGVLANSGDLALHDYVVAHFVYVMSLGAVFALFACFYVWVGKIFGR 405
QY      438  AFPEMAARLHEWTFEIGANVTFFPQHLLGROGMPPRYIDYPEAFALMNSVSGAFLFA 497
Dd      406  TYPELIGQIHWFITFEFGVNLTFEPNHLFGLSGMPRIPIDYPRAYAGWMLSSFGSIVSY 465
QY      498  SFLFVITVYFVTVLVAGR--RETRPDPWG--EPADLTLEWTLSPRAHHTFEETLP 546

```

Db	466	GICCFVYVVTITLSSGKNKRCAPSWAWEAENSTILEMWAQSPAPHTITGELP	517
RESULT	8		
ID	COX1_RICPR	STANDARD:	PRT: 534 AA.
AC	054069;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME Aa3 SUBUNIT 1).		
DE	CTAD OR COXA OR RP405.		
OS	Rickettsia prowazekii.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;		
OC	Rickettsiaceae; Rickettsiase; Rickettsia.		
OX	NCBI_TaxID=782;		
RN			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MADRID E;		
RC	Sicheritz T., Kurland C.G., Andersson S.G.E.;		
RT	"The bacterial origin of mitochondria inferred from a phylogenetic analysis of the cytochrome b and cytochrome c oxidase i genes."		
RL	Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MADRID E;		
RX	MEDLINE=99039499; PubMed=9823893;		
RA	Andersson S.G.E., Zomorodipour A., Andersson J.O.,		
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,		
RA	Eriksson A.-S., Winkler H.H., Kurland C.G.;		
RT	"The genome sequence of Rickettsia prowazekii and the origin of mitochondria."		
RT	Nature 396:133-140(1998).		
RL			
CC	-I- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).		
CC	-I- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4 FERRICYTOCHROME C.		
CC	-I- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.		
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
CC	-I- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL; Y13855; CAA74167.1; -		
DR	EMBL; AJ235271; CAA14862.1; -		
DR	InterPro: IPR000883; COX1.		
DR	Pfam: PF00115; COX1.1.		
DR	PRINTS; PRO1165; CYCOXIDASE1.		
DR	ProSITE; PS00077; COX1.1.		
KW	Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain; Complete proteome.		
KW	TRANSMEM 35 55	POTENTIAL.	
FT	TRANSMEM 76 96	POTENTIAL.	
FT	TRANSMEM 97 117	POTENTIAL.	
FT	TRANSMEM 120 140	POTENTIAL.	
FT	TRANSMEM 165 185	POTENTIAL.	
FT	TRANSMEM 202 222	POTENTIAL.	
FT	TRANSMEM 254 274	POTENTIAL.	
FT	TRANSMEM 286 306	POTENTIAL.	
FT	TRANSMEM 320 340	POTENTIAL.	





DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 CC CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COXI OR COXI.  
 OS Zea mays (Maize).  
 OG Mitochondrion.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 CC Panicoideae; Andropogoneae; Zea.  
 CC NCBI\_TaxId=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isaac P.G., Jones V.P., Leaver C.J.;  
 RT "The maize cytochrome c oxidase subunit I gene: sequence, expression  
 RT and rearrangement in cytoplasmic male sterile plants.";  
 RT EMBO J. 4:1617-1623(1985).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
 CC 4 FERRICYTOCHROME C.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X02660; CAA26496.1; -.  
 CC PIR: A22840; ODZM1.  
 CC HSSP: P00396; LOCC.  
 CC MaizeDB: 69214; -.  
 CC -----  
 CC DR Mendel: 2201; Zeama:coxi.1.  
 CC InterPro: IPR000883; COXI.  
 CC DR Pfam: PF00115; COXI.1.  
 CC DR PRINTS: PRO1165; CYCOXIDASE1.  
 CC DR PROSITE: PS00077; COXI.1.  
 CC KM Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 CC Respiratory chain; Inner membrane.  
 CC FT METAL 64 64 IRON (HEME A) (PROBABLE).  
 CC FT METAL 243 243 COPPER B (PROBABLE).  
 CC FT METAL 247 247 COPPER B (PROBABLE).  
 CC FT METAL 292 292 COPPER B (PROBABLE).  
 CC FT METAL 293 293 COPPER B (PROBABLE).  
 CC FT METAL 378 378 IRON (HEME A3) (PROBABLE).  
 CC FT METAL 380 380 IRON (HEME A) (PROBABLE).  
 CC SEQUENCE 528 AA; 58257 MW; 75FB95568E9E7 CRC64;  
 SQ  
 Query Match 57.7%; Score 1730.5; DB 1; Length 528;  
 Best Local Similarity 58.5%; Pred No. 1e-93;  
 Matches 311; Conservative 88; Mismatches 108; Indels 25; Gaps 5;  
 Oy 20 RWEFSTNHRDGLILYLAAGVVGFIISLFTYVMKRLMDPGVQVMCLGEGARLADASQTC 79  
 Db 6 RWEFSTNHRDGLILYLAAGVVGFIISLFTYVMKRLMDPGVQVMCLGEGARLADASQTC 79  
 Oy 80 TANGHLMNVAWYHGLMFEFGVIGALPLFGFNGYMLPDLIGAPDAFPPANNLSFWLFTA 139  
 Db 51 GGHQHOYNNLITFAHAFIMTFEVMFPMIGFGNMFVPLIGAPDAFPPANNLSFWLFTA 110  
 Oy 140 GTAMGVASLFPARGDGLSGVGVWVLPPLS--TREAGYSMDLAFAVHLSSSSIMGAI 197  
 Db 111 SLULLLSALV-----EVSGGTGWIVYPPPLSGITSHSGAVDLATFSLHSGVSSILGSI 165

Oy 198 MNTTFLNMRAPGWTLLKRVPLFSWISIFITAMLLILALPVLAGAITMLTDNRNGTFFNR 257  
 Db 166 NMTTFLNMRAPGWTLLKRVPLFSWISIFITAMLLILALPVLAGAITMLTDNRNGTFFNR 255  
 Oy 258 AGGDDPLLYOHIIMFEGHPEVYIIILDFGSLHVSSTSKKRVFGYLVYVMAVAGVL 317  
 Db 226 AGGDDPLLYOHIIMFEGHPEVYIIILDFGSLHVSSTSKKRVFGYLVYVMAVAGVL 285  
 Oy 318 GFVVAHMHMTYVGLSLQOOSYFMATWVIAVPGIKFISMTATMGSGVFEKSPMLAFC 377  
 Db 286 GFVVAHMHMTYVGLSLQOOSYFMATWVIAVPGIKFISMTATMGSGVFEKSPMLAFC 345  
 Oy 378 FMEPLTGVGTGIVLAAGLDRAVHDYVYVAHFRHYMSLGAIFALFAGIYFMPKFSGR 437  
 Db 346 FMEPLTGVGTGIVLAAGLDRAVHDYVYVAHFRHYMSLGAIFALFAGIYFMPKFSGR 405  
 Oy 438 AFPEMAKLHMTFTFICANVTFFPHGLGKQMPRRIDYDPEAFALNNKSSGAFALFA 497  
 Db 406 TYPFTLQIHFWITFFGVNLIFFPMHFLGSGMPRRIDYDPEAFALNNKSSGAFALFA 465  
 Oy 498 SLEFVIEVYTVLAGR--RETRPNMG--EFADPLEWTLSPPPAHTFTLP 546  
 Db 466 GIRPEFVVAITSSGKKNKRAESPAVEONPTTLEMLVOSPPAFHTFGELP 517  
 RESULT 12  
 ID COXI\_SORBI STANDARD; PRT; 530 AA.  
 AC P05502;  
 DT 01-NOV-1998 (Rel. 09, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COXI OR COXI.  
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 OG Mitochondrion.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 CC Panicoideae; Andropogoneae; Sorghum.  
 CC NCBI\_TaxId=4558;  
 RN [1]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RX MEDLINE=87051727; PubMed=3022938.  
 RA Bailey-Serres J., Hanson D.K., Fox T.D., Leaver C.J.;  
 RT "Mitochondrial genome rearrangement leads to extension and relocation  
 RT of the cytochrome c oxidase subunit I gene in sorghum.";  
 RL Cell 47:567-576(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hanson D.K., Bailey-Serres J., Leaver C.J.;  
 RL Submitted (May-1987) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
 CC 4 FERRICYTOCHROME C.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----





DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1.  
 OS Chondrus crispus (Carragheen).  
 OG Mitochondrion.  
 OC Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartineaceae;  
 OC Chondrus.  
 CC NCBI\_TaxID=2769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Apices;  
 RX MEDLINE=95341681; PubMed=7616569;  
 RA Leblanc C., Boyen C., Richard O., Bonnard G., Grienenberger J.M.,  
 RA Klarg B.;  
 RT "Complete sequence of the mitochondrial DNA of the rhodophyte  
 RT Chondrus crispus (Gigartinales). Gene content and genome  
 RT organization.";  
 RL J. Mol. Biol. 250:484-495(1995).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
 CC 4 FERROCYTOCHROME C.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; 247547; CAAB7603.1; -;  
 DR HSSP; P00396; 10CC.  
 DR Mendel; 7724; CHOCr;cox1.1.  
 DR InterPro; IPR000883; COX1.  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASE1.  
 DR PROSITE; PS00077; COX1; 1.  
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 KW Respiratory chain; Inner membrane.  
 FT METAL 69 69 IRON (HEME A) (PROBABLE).  
 FT METAL 248 248 COPPER B (PROBABLE).  
 FT METAL 252 252 COPPER B (PROBABLE).  
 FT METAL 297 297 COPPER B (PROBABLE).  
 FT METAL 298 298 COPPER B (PROBABLE).  
 FT METAL 383 383 IRON (HEME A3) (PROBABLE).  
 FT METAL 385 385 IRON (HEME A3) (PROBABLE).  
 SEQUENCE 532 AA; 59290 MW; ADF18B4B35BA636F CRC64;

Query Match 55.9%; Score 1677.5; DB 1; Length 532;  
 Best Local Similarity 56.1%; Pred. No. 1.2e-90;  
 Matches 306; Conservative 94; Mismatches 110; Indels 35; Gaps 7;

QY 15 QGFEF-----RWMFNTNHDIGLTVYAGVGFISLVTVYRLMDPGVOYMCLEGAR 70  
 DB 2 QSEFTQWISRWTFSTNHRDGLTVLFGAFSGVLGCGWSMILRMELAP----- 50  
 QY 71 LIADASQTCFANGHLAMNVYTHGILMFVCGIPALFGGFGNYLMPLOIGADPMAPRRN 130  
 DB 51 -----SNHLLGHNQIYLVLIATHAFMLTFPMVPMVIGFGNWLVPIMIGSPDMAFPRLN 106

QY 131 NLSFWLFIAGTAGVAVSLFAPGCGDGLGSGVGVLYPLST--REAGYSMDLAFVHUS 188  
 DB 107 NISFWLLPISLCLLMSALY-----EVGVTGWTVPPLSSIQSHSGAVDLAFESLHS 161  
 QY 189 GASSIMGAIMNTTFFLNMRAPGKTLKVPDLFSWSIFITAMLILALPVLGATMTLTD 248  
 DB 162 GASSILGAVNFISTLNMRSPPGSMRIFLPFWSILVTFALLAVPVLGATMTLTD 221  
 QY 249 NGCTTFENAGGDDPLLYOHILMFSGHPEVYIILDFGLISHWVSFESKRPVGLPW 308  
 DB 222 NPTSFEDASGGDPLLYOHLFWFPGHPEVYIILDFGFMISHWVSFESKRPVGYIGW 281  
 QY 309 YAMVAIGVGFVYVAHMYTVGSLTQGSYFMLATVIAVPTGIRKFSMTATWKGVSVE 368  
 DB 282 YAMVSGVIGFVYVAHMYTVGSLTQGSYFMLATVIAVPTGIRKFSMTATWKGVSVE 341  
 QY 369 KSPMLAFSGFMELFTVGVGTGLVLAQGLDRAVHDYVYVAHMYVSLCAIFAFAGTY 428  
 DB 342 KTPMLFAIFGIFELFTIGLGLVLAQGLDRAVHDYVYVAHMYVSLCAIFAFAGTY 401  
 QY 429 FYMPKFGSAPFEMAKLHPWTFEIGANTFFPQHLGQGMRRIDYDEAFALNNKVS 488  
 DB 402 YWFGKITGLQYPTLGLQHFWSFTFIGNLTFMPHFLGLAGMPRLDPDVAAGNLLA 461  
 QY 489 SYGAFLAFAFSLFFIYFYVTLVAGRRTRPNPWCSEF-----ADTLFWTLPSPPAHT 541  
 DB 462 SYGSLTALSTLFFFIYFYVTLVAGRRTRPNPWCSEF-----ADTLFWTLPSPPAHT 519  
 QY 542 FETLP 546  
 DB 520 FEEMP 524

Search completed: December 22, 2001, 11:08:07  
 Job time: 336 sec





[illegible]

QY	122	PDAFPPNNLWSPWLFAGTAMCVASLFPADGGDGLSGVGVLPPLST-REAGYSMDL	180
Db	121	PDAFPPNNLWSPWLFAGTAMCVASLFPADGGDGLSGVGVLPPLST-REAGYSMDL	179
QY	181	AIFAVHLSGASSIMGAINMTTFFLNRRABGMLLHKVPLFWSIFITWMLLALPVLAGA	240
Db	180	AIFAVHLSGASSIMGAINMTTFFLNRRABGMLLHKVPLFWSIFITWMLLALPVLAGA	239
QY	241	ITMLLTDRNGSTTFPPNADGGDRPLVONHLMFEGHREYVLIILPBEGLISHVVSFESKRP	300
Db	240	ITMLLTDRNGSTTFPPNADGGDRPLVONHLMFEGHREYVLIILPBEGLISHVVSFESKRP	299
QY	301	VFGYLPVYVMAVIGVLEGVVMAHHNYTGMSLTDOOSYMLATWVIAVPTGIKIFSWIAT	360
Db	300	VFGYLPVYVMAVIGVLEGVVMAHHNYTGMSLTDOOSYMLATWVIAVPTGIKIFSWIAT	359
QY	361	MGGSGVEFKSPMLABCFMMLLFTVGGVGTGLVLAQALDPLAAYDTPYVVVAHHFHYVMSGAI	420
Db	360	MGGSGVEFKSPMLABCFMMLLFTVGGVGTGLVLAQALDPLAAYDTPYVVVAHHFHYVMSGAI	419
QY	421	FAIFAGSYTFWPKFSGRAPFEMAALHMTFFIGANWTFPPOHFLSGROGMRPYIDYEA	480
Db	420	FAIFAGSYTFWPKFSGRAPFEMAALHMTFFIGANWTFPPOHFLSGROGMRPYIDYEA	479
QY	481	FALNNKVVSSGCAIAPASLFFLYIVIVYVTLVAGRRTRRPNPGEFADLTLEWLPSPRPAH	540
Db	480	FALNNKVVSSGCAIAPASLFFLYIVIVYVTLVAGRRTRRPNPGEFADLTLEWLPSPRPAH	538
QY	541	TFETLPKRSDMDKH	554
Db	539	QFSEPPYVTKGDH	552

Result	ID	PRELIMINARY:	PRT:	539 AA.
059631				
AC	Q59631.			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)			
DE	CYTOCHROME-C OXIDASE SUBUNIT I (EC 1.9.3.1) (CYTOCHROME OXIDASE)			
DE	(CYTOCHROME A(3)) (CYTOCHROME AA(3)).			
GN	COXB.			
OS	Nitrobacter winogradskyi (Nitrobacter agilis).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Bradyrhizobium group; Nitroacteraceae; Nitroacter.			
OX	NCBI_TaxId=913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 14123;			
RA	Berben G.;			
RL	Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCITOCROME C + O(2) = 4 FERRICITOCROME			
CC	C + 2 H(2)O.			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCITOCROME C + O(2) = 2 H(2)O + 4			
CC	FERRICITOCROME C.			
CC	-1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).			
CC	-1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.			
DR	EMBL; X89566; CA61744.1; -.			
DR	HSSP; P98002; IARI.			
DR	InterPro; IPR000883; COX1.			
DR	Pfam; PF00115; COX1; 1.			
DR	PRINTS; P001165; CYCOXIDASE1.			
DR	PROSITE; PS00077; COX1; 1.			
KW	Copper; Heme; Oxidoreductase; Respiratory chain; Transmembrane.			
SQ	SEQUENCE 539 AA: 59145 MW; 2639148DDAC23628 CRC64;			

Query Match	61.3%;	Score 1838;	DB 2;	Length 539;
Best Local Similarity	61.8%;	Pred. No. 4.6e-93;		

Matches 345; Conservative 74; Mismatches 107; Indels 32; Gaps 7;

[illegible]

```

RESULT      3
ID           036520
AC           036520      PRELIMINARY;      PRT;      523 AA.
DC           036520:
DT           01-NOV-1996 (TREMBLrel. 01, Created)
DT           01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT           01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE           CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN           COXI OR COXI.
OS           PLATYMONAS SUBCORDIFORMIS.
OC           Mitochondrion.
OC           Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Platymonas.
OX           NCBI_TaxID=3161;
RN           [1]
RP           SEQUENCE FROM N.A.
RL           Kessler U.;
RL           Thesis (1995), Institute for Plant Physiology,
RL           Justus Liebig University, Giessen, Germany.
RN           [2]
RP           SEQUENCE FROM N.A.
RX           MEDLINE=9614517; PubMed=8555450;
RA           Kessler U.; Zetsche K.;
RT           "Physical map and gene organization of the mitochondrial genome from
RT           the unicellular green alga Platymonas (Tetraselmis) subcordiformis
RT           (Prasinophyceae).";
RL           Plant Mol. Biol. 29:1081-1086(1995).
CC           -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC           CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-1
CC           3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
CC           CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN

```

CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
CC FERRICYTOCHROME C.  
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC EMBL: 247795; CAA87753.1; .  
DR HSP: P98002; IRL1.  
DR Mendel: 21543; Plasu;cox1;21543.  
DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS: P001165; CYCOXIDASE1.  
DR PROSITE: PS00077; COX1; 1.  
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
SQ SEQUENCE 523 AA; 57669 MW; 3E5B17B6E749085C CRC64;

Query Match 59.1%; Score 1773; DB 8; Length 523;  
Best Local Similarity 60.1%; Pred. No. 1.6e-89;  
Matches 322; Conservative 82; Mismatches 104; Indels 28; Gaps 6;

QY 17 FETRMFSTNKHDKDGLLYLVAGVGFISVLFYVYMRLELMDPGVOYMCLEGARLIADAS 76  
DB 4 FAQRMWSTNKHDKDGLLYLVAGVGFISVLFYVYMRLELMDPGVOYMCLEGARLIADAS 48  
QY 77 QCTTANGHLNVAWVYHGLIMFVGVIPALFEGFGNYLMPLOIGADMAPRPMNLSFWL 136  
DB 49 QVLGNHOLYVNYITAHAFLEIFMVPBALGEGNMFVPMIGADMAPRPMNLSFWL 108  
QY 137 FIAGTANGVASTLPARGDGOIGSGGVNLYPRPST--REAGYSMDLIPAVHLSGASSIM 194  
DB 109 LPPSLLLSLLSALV-----ELGAGTGWTVYPPSLSIASHSGASVDLAFSLHSIGASSIL 163  
QY 195 GAINMTTFLNMRAPGMLTAKYVLFWSIFITAWLITLALPYLAGATMLLDRNFGTTE 254  
DB 164 GALNFITTLNMRPGMTNHRPLFVWSVLTATLALLSLPYLAGATMLLDRNFGTTE 223  
QY 255 FNPAGGDPILYOHILWFEGHPEVYIIILPGFGLISHVSTSKRPVGYLPMYANVAI 314  
DB 224 FDPAGGDPILYOHILWFEGHPEVYIIILPAFGILSHVSTSKRPVGYLPMYANVAI 283  
QY 315 GVLGFVVAHHNMYTVGMSLTOOSYFMLATVAVPTGKIFSMATMGGSVERKSPMLM 374  
DB 284 GILGFIVMAHHMYTVGDLIDITRAITPAATMIIVPTGKIFSMATMGGSVERKSPMLF 343  
QY 375 AFGMFLETVGGVTVGLAQAAGLDRAVHDTYVVVAHFHYVMSLGAIFAFAGIYFMPKF 434  
DB 344 AIGFLFLTVGGVTVGLVANSGLDIALHDYVVVAHFHYVMSLGAIFAFAGIYFMPKF 403  
QY 435 SGRAPFPMMAKLHWTFITGANTFFPOHFLRGOMPRRYIDYPAEALNNKVSYGAF 494  
DB 404 CGLOYSSEFLQIHFWFEGHPEVYIIILPGFGLISHVSTSKRPVGYLPMYANVAI 463  
QY 495 AFAFSEFLFVTVGLVAGRRETRPNF---GEFADTLMTLPPSPAPRTFELP 546  
DB 464 SVLGALFEFFVY-VYDITLSGERCKPKMVEVPGTSA-TLEMLTPSPAPYHFFEEVP 517

RESULT 4  
QYTCAG PRELIMINARY; PRT: 525 AA.  
AC QYTCAG;  
DT 01-MAY-2000 (TREMBLREL. 13, Created)  
DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
GN COX1.  
OS Nephroselmis olivacea.  
OC Mitochondrion.

OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;  
OC Chlorodendraceae; Chlorodendraceae; Nephroselmis.  
OX NCB1\_TaxID=31312;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NIES-484;  
RX MEDLINE-99418884; PubMed-10488238;  
RA Turmel M., Lemieux C., Burger G., Lang B.F., Otis C., Plante I.,  
RA Gray M.W.;  
RT "The complete mitochondrial DNA sequences of Nephroselmis olivacea and  
RT Pedionomonas minor: two radically different evolutionary patterns  
RT within the green algae."  
RL Plant cell 11:1717-1729(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NIES-484;  
RA Turmel M., Lemieux C., Burger G., Lang B.F., Otis C., Plante I.,  
RA Gray M.W.;  
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
CC FERRICYTOCHROME C.  
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL: AF101368; AAF03191.1; .  
DR HSP: P18401; IFTT.  
DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS: P001165; CYCOXIDASE1.  
DR PROSITE: PS00077; COX1; 1.  
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
SQ SEQUENCE 525 AA; 58099 MW; 04293667EBAD4AA CRC64;

Query Match 58.5%; Score 1756.5; DB 8; Length 525;  
Best Local Similarity 58.2%; Pred. No. 1.2e-88;  
Matches 312; Conservative 90; Mismatches 107; Indels 27; Gaps 5;

QY 17 FETRMFSTNKHDKDGLLYLVAGVGFISVLFYVYMRLELMDPGVOYMCLEGARLIADAS 76  
DB 4 FVQRMWSTNKHDKDGLLYLVAGVGFISVLFYVYMRLELMDPGVOYMCLEGARLIADAS 48  
QY 77 QCTTANGHLNVAWVYHGLIMFVGVIPALFEGFGNYLMPLOIGADMAPRPMNLSFWL 136  
DB 49 QVLGNHOLYVNYITAHAFLEIFMVPBALGEGNMFVPMIGADMAPRPMNLSFWL 108  
QY 137 FIAGTANGVASTLPARGDGOIGSGGVNLYPRPST--AGSMDLIPAVHLSGASSIM 194  
DB 109 LPPSLLLSLLSALV-----EVGAGTGWTVYPPSLSIASHSGASVDLAFSLHSIGASSIL 163  
QY 195 GAINMTTFLNMRAPGMLTAKYVLFWSIFITAWLITLALPYLAGATMLLDRNFGTTE 254  
DB 164 GALNFITTLNMRPGMTNHRPLFVWSVLTATLALLSLPYLAGATMLLDRNFGTTE 223  
QY 255 FNPAGGDPILYOHILWFEGHPEVYIIILPGFGLISHVSTSKRPVGYLPMYANVAI 314  
DB 224 FDPAGGDPILYOHILWFEGHPEVYIIILPAFGILSHVSTSKRPVGYLPMYANVAI 283  
QY 315 GVLGFVVAHHNMYTVGMSLTOOSYFMLATVAVPTGKIFSMATMGGSVERKSPMLM 374  
DB 284 GILGFIVMAHHMYTVGDLIDITRAITPAATMIIVPTGKIFSMATMGGSVERKSPMLF 343  
QY 375 AFGMFLETVGGVTVGLAQAAGLDRAVHDTYVVVAHFHYVMSLGAIFAFAGIYFMPKF 434

Db 344 AVGFLEFTIGFTGILLNSGDLIALHDYVVVGHFHVLSMAGVFGNFAFYWGK1 403

Qy 435 SGAPAFEMNAKHEWTFEFGANVTEPPQHELRQGMREYIDYPEAFAMNVSYSGL 494

Db 404 TGLQYETLGLQHEWTFEFGVNFHFLGLAGMPRIIDYPDAYAGMNAIASGSL 463

Qy 495 AFSPFEEVIEVYFVAVGRRETRPMPG----EFADLEWTLPSPPHPTETLP 546

Db 464 SILGALFFEVV-VYATLTGNEKGNMNPADKRDVASTLEMMVSGSPAPHTHO1P 518

RESULT 5

063228 PRELIMINARY: PRT: 528 AA.

AC 063228:

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).

GN COX1.

OS Populus tremuloides (Quaking aspen).

OC Mitochondrion.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids I; Malpighiales; Salicaceae; Populus.

OX NCBI\_TaxID=3693;

RN [1]

RP SEQUENCE FROM N.A.

RP TISSUE-LEAF;

RC MEDLINE=97303197; PubMed=9159140;

RA Laroche J., Li P., Maggia L., Bousquet J.;

RT Molecular evolution of angiosperm mitochondrial introns and exons.";

Proc. Natl. Acad. Sci. U.S.A. 94:5722-5727(1997).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: U77623; AAC98473.1; -

DR HSSP: P98002; IAR1.

DR Mendel: 23876; Poplm:cox1;23876.

DR InterPro: IPR000883; COX1.

DR Pfam: PF00115; COX1; 1.

DR PROSITE: PS00077; COX1; 1.

KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane.

FT NON\_TER 528 528

FT SEQUENCE 528 AA; 57993 MW; 46600F70CD4D5047 CRC64;

Query Match 58.4%; Score 1751.5; DB 8; Length 528;

Best Local Similarity 59.0%; Pred. No. 2.3e-88;

Matches 314; Conservative 89; Mismatches 104; Indels 25; Gaps 5;

Qy 20 RWMSTNHKDIGILLYVAGVGVFISLFTVYMRLEMDPGVQVMCLLEGARLIDASQTC 79

Db 6 RMLFSTNHKDIGILLYVAGVGVFISLFTVYMRLEMDPGVQVMCLLEGARLIDASQTC 79

Qy 80 TANGHLMNVNVTYHGLIMEFVSGIPALFGEGVNYLMDLGADMAFPKRNNSFWLFTA 139

Db 51 GGNHQLYVNLITAHAFIMLFPMVPMIGGFGWVFPIILIGADMAFPRLNNSFWLPP 110

Qy 140 GTAMGVAFLFAPGGDGLSGVGVWVLYPLPS--TRRAGSMDLATFAVHLSGSSIMGAI 197

Db 111 SLTLLLSALV-----EVSGTGWTYVPLPSLSSHGAVDLAIFSLHLSGVSTLGS1 165

Qy 198 MNTTFLMRAFGMTLHKLPLEFSISFTAWLILALPVLAGATMLTDNRNGTTFNP 257

Db 166 NPTITIFMRGRGNTMHRLEPLFVMSVLVATFALLSLSPVLGATMLTDNRNTTFEED 225

Qy 258 AGGSDPLLYQHILWFEGHPEVYIILDFEGITSHVSTFSKRPVEGYLPMVYANVAIGV 317

Db 226 AGGGDPLLYQHILWFEGHPEVYIILDFEGITSHVSTFSKRPVEGYLPMVYANVAIGV 285

Qy 318 GFVVAHNMVTVGSLTQOQSYEMLATVIANPQIKLFSWTATMGGSVFEKSPMLAFG 377

Db 286 GFLVANHMFVGLVDVTRAFYATMTIAVPTGKITFSWTATMGGSIDQKTPMLPVG 345

Qy 378 FMFLTVAGVGTIVLAQGLDRAVHDYVVVVAHFHYMSGALFAIRAGIYFVMPKFSGR 437

Db 346 FIFLETLGLGIYLANGLDIALHDYVVVGHFHVLSMAGVFGNFAFYWGK1TGR 405

Qy 438 APPEMAKHEWTFEFGANVTEPPQHELRQGMREYIDYPEAFAMNVSYSGLAFA 497

Db 406 TYPEPLGKHFWITFEFGVNFHFLGLAGMPRIIDYPDAYAGMNAIASGSL 465

Qy 498 SPLFFIVFVYTVLAG-RRETRPMPG--EFADLEWTLPSPPHPTETLP 546

Db 466 GICCFVYVVTTLSSGNONKCAPSPWALDONSTLEMMVSGSPAPHTFGELP 517

RESULT 6

037616 PRELIMINARY: PRT: 515 AA.

AC 037616:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).

GN COX1.

OS Prototheca wickerhamii.

OC Mitochondrion.

OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorrellales;

OC Chlorrellaceae; Prototheca.

OX NCBI\_TaxID=3111;

RN [1]

RP SEQUENCE FROM N.A.

RP STRAIN=263-11;

RC MEDLINE=94180393; PubMed=8135522;

RA Wolff G., Plante I., Lang B.F., Kueck U., Burger G.;

RT "Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii. Gene content and genome organization.";

RL J. Mol. Biol. 237:75-86(1994).

RN [2]

RP SEQUENCE FROM N.A.

RP STRAIN=263-11;

RC MEDLINE=94180393; PubMed=8135522;

RA Wolff G., Burger G., Lang B., Kueck U.;

RT "Mitochondrial genes in the colourless alga Prototheca wickerhamii resemble plant genes in their exons but fungal genes in their introns.";

RL Nucleic Acids Res. 21:719-726(1993).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: U02970; A012634.1; -

DR HSSP: P98002; IAR1.

DR Mendel: 21560; Protol:cox1;21560.

DR InterPro: IPR000883; COX1.  
 DR Pfam: PF00115; COX1.1.  
 DR PRINTS: PR01165; CYCOXIDASE1.  
 DR PROSITE: PS00077; COX1.1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 SO SEQUENCE 515 AA; 57036 MW; 2F6D1AD87E240B CRC64;

Query Match 58.2%; Score 1745; DB 8; Length 515;  
 Best Local Similarity 58.2%; Pred. No. 5.1e-88;  
 Matches 310; Conservative 92; Mismatches 103; Indels 28; Gaps 6;

QY 19 TRWENSTNKKDIGLLYLVAAGVGFISLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 78  
 DB 3 TRWLYSTNKKDIGLLYLVAAGVGFISLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 47  
 QY 79 CTANGLMNNVWVYHGLIMFVFGVIFALFGFGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 138  
 DB 48 LKGNHQLVNLITLHAFLMIFPMPLMGFGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 107  
 QY 139 AGTAMGVASLFAFGDGLSGVGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 196  
 DB 108 PSLILVSSALV-----EVGAGTGWYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 162  
 QY 197 INMTTFLNRRAPMTLHKYVPLFSGVIFITAMLLALPVLGATITMLLDRNGTTFEN 256  
 DB 163 INMTTFLNRRAPMTLHKYVPLFSGVIFITAMLLALPVLGATITMLLDRNGTTFEN 222  
 QY 257 PGGGDPILYOHILMFPGHREYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 316  
 DB 223 PGGGDPILYOHILMFPGHREYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 282  
 QY 317 LGFVMAHMYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 376  
 DB 283 LGFVMAHMYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 342  
 QY 377 GFMFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 436  
 DB 343 GFMFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 402  
 QY 437 RAPFEMAKLHEFTFGAVTEFFQHLRQGMRRYIDYPAFAIMNNVSSGATLAF 496  
 DB 403 LQYFETLQHLHFMFLGAVTEFFQHLRQGMRRYIDYPAFAIMNNVSSGATLAF 462  
 QY 497 ASLFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 545  
 DB 463 ASLFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 513

RESULT 7  
 ID 035812 PRELIMINARY; PRT; 632 AA.  
 AC 035812;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1.  
 OS Sorghum sp. (sorghum).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Sorghum.  
 OC NCBI\_taxid=4559;  
 RN [1]  
 RP SEQUENCE OF 502-632 FROM N.A.  
 RC STRAIN=MUTANT 9E;  
 RX MEDLINE=8705127; PubMed=3022938;  
 RA Bailey-Serres J., Hanson D.K., Fox T.D., Leaver C.J.;  
 RT Mitochondrial genome rearrangement leads to extension and relocation  
 RL of the cytochrome c oxidase subunit I gene in sorghum. "Cell 47:567-576(1986).

RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MUTANT 9E;  
 RA Hanson D.K., Bailey-Serres J., Leaver C.J.;  
 RL Submitted (JUN-1987) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE  
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
 CC FERROCYTOCHROME C.  
 CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: M14454; AAA70319.1; -.  
 DR HSSP: P98002; IAR1.  
 DR Mendel: 2179; Sorgh; COX1; 2179.  
 DR InterPro: IPR000883; COX1.  
 DR Pfam: PF00115; COX1.1.  
 DR PRINTS: PR01165; CYCOXIDASE1.  
 DR PROSITE: PS00077; COX1.1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 SO SEQUENCE 632 AA; 70204 MW; 85575B1A84603D80 CRC64;

Query Match 57.7%; Score 1730.5; DB 8; Length 632;  
 Best Local Similarity 58.5%; Pred. No. 3.9e-87;  
 Matches 311; Conservative 88; Mismatches 108; Indels 25; Gaps 5;

QY 20 RWPSTNKKDIGLLYLVAAGVGFISLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 79  
 DB 6 RWPSTNKKDIGLLYLVAAGVGFISLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 50  
 QY 80 TANGHLMNNVWVYHGLIMFVFGVIFALFGFGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 139  
 DB 51 TANGHLMNNVWVYHGLIMFVFGVIFALFGFGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 110  
 QY 140 GTAMGVASLFAFGDGLSGVGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 197  
 DB 111 GTAMGVASLFAFGDGLSGVGVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 165  
 QY 198 NMTTTFLNRRAPMTLHKYVPLFSGVIFITAMLLALPVLGATITMLLDRNGTTFEN 257  
 DB 166 NMTTTFLNRRAPMTLHKYVPLFSGVIFITAMLLALPVLGATITMLLDRNGTTFEN 225  
 QY 258 PGGGDPILYOHILMFPGHREYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 317  
 DB 226 PGGGDPILYOHILMFPGHREYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 285  
 QY 318 LGFVMAHMYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 377  
 DB 286 LGFVMAHMYVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 345  
 QY 377 GFMFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 437  
 DB 346 GFMFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 405  
 QY 438 RAPFEMAKLHEFTFGAVTEFFQHLRQGMRRYIDYPAFAIMNNVSSGATLAF 497  
 DB 406 RAPFEMAKLHEFTFGAVTEFFQHLRQGMRRYIDYPAFAIMNNVSSGATLAF 465  
 QY 498 ASLFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 546  
 DB 466 ASLFLVYVYKRLKELMDPVGQVYMLLEGARLIADASQT 517

RESULT 8  
 ID 035812 PRELIMINARY; PRT; 632 AA.  
 AC 035812;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1.  
 OS Sorghum sp. (sorghum).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Sorghum.  
 OC NCBI\_taxid=4559;  
 RN [1]  
 RP SEQUENCE OF 502-632 FROM N.A.  
 RC STRAIN=MUTANT 9E;  
 RX MEDLINE=8705127; PubMed=3022938;  
 RA Bailey-Serres J., Hanson D.K., Fox T.D., Leaver C.J.;  
 RT Mitochondrial genome rearrangement leads to extension and relocation  
 RL of the cytochrome c oxidase subunit I gene in sorghum. "Cell 47:567-576(1986).

ID 096875 PRELIMINARY; PRT: 532 AA.  
 AC 096875:  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1.  
 OS Rhodomonas salina.  
 OC Mitochondrion.  
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Rhodomonas.  
 OC NCBI\_TaxID=52970;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Burger G., Lang B.F., Maier U.G., McPadden G.I., Gray M.W.;  
 RT "Algae with secondary chloroplasts have mitochondria that originate from the host."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4 FERROCYTOCHROME C.  
 CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL: AF288090; AAG17662.1; -.  
 DR InterPro: IPR000883; COX1.  
 DR Pfam: PF00115; COX1.1.  
 DR PRINTS: PR01165; CYCOXIDASEL.  
 DR PROSITE: PS00077; COX1.1.  
 DR Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 DR Respiratory chain; Transmembrane.  
 KM Respirocytochrome c.  
 KW SEQUENCE 532 AA; 59152 MW; 64E39A3CD29CEFE4 CRC64;  
 SQ

Query Match 57.4%; Score 1720.5; DB 8; Length 532;  
 Best Local Similarity 57.5%; Pred. No. 1.2e-86;  
 Matches 311; Conservative 90; Mismatches 107; Indels 33; Gaps 6;

17 FFTRMSTNHRDGLLYLAAGVGVFSVLTYYMRLEMDPGVQYMCLEGARLIADAS 76  
 3 FTRMFSTNHRDGLLYLAAGVGVFSVLTYYMRLEMDPGVQYMCLEGARLIADAS 50  
 77 QCTANGHLMNVVYTHGLMFEFGVIRALRGFGNYLPLDQAGDAFPRMNNLSFWL 136  
 51 ---GGNHOLYNYLVIGHAFIMFEFVMPALIGYGNELPVMIGAVDAFPRMNNVSEWL 107  
 137 FLAGTMGVASLEFAPGCGDQLSGVGVNLYPLSTREA--GSMDLAIFAVHLSGASSIM 194  
 108 LPPALLLISSTLTGEG----GAGTGTYYPPPLSSVEGSPSAIDLDGIFSLHVAAGASSIL 162  
 195 GAINMTTELNMKRAPGMLHKVPLFSMSIFITAMILLALPVLAGATMILLDRNFGTTE 254  
 163 GAINMTTELNMKRAPGMLHKVPLFSMSIFITAMILLALPVLAGATMILLDRNFGTTE 222  
 255 FNPAGGDDPLVQHLIMFEGHPEVYIIIPRGIIISVVSTSKRVEGYLPMVYAVMAI 314  
 223 FNPAGGDDPLVQHLIMFEGHPEVYIIIPRGIIISVVSTSKRVEGYLPMVYAVMAI 282  
 315 GVGFEVVMVNHVYVMSLTQOQSYFMLATMVAIVPTGKIFSMIAIMMGSGVEFSPML 374  
 283 GLGFEVVMVNHVYVMSLTQOQSYFMLATMVAIVPTGKIFSMIAIMMGSGVEFSPML 342  
 375 AFGEFELFTVGVGTIVLAQAGLDRAVHDTYVVVAHFHYVMSIGALFALFAGIYMPKE 434  
 343 AVGEFELFTVGVGTIVLAQAGLDRAVHDTYVVVAHFHYVMSIGALFALFAGIYMPKE 402  
 435 SGRAPEMAKHLFWTFEFGANVTFFQHLGQGNPRKIDYIEPAFLMNNVSSIGA-F 493

DB 403 TGFQYRNLGVHIFWCFVGVNLTFFQHFHGLAGMPRRIPDYSYAMNMSISGSYF 462  
 QY 494 LAFASLEFIVIVYTVIAGRRRETRPMV-----GEPAOTLEMTLSPPAHFEETL 545  
 DB 463 SVFIFLEFVLT--VETLTMEQCEVPMKFSMEDVKNDREYLEWLVGSPPAHFEETL 520  
 QY 546 P 546  
 DB 521 P 521

RESULT 9  
 ID 099988 PRELIMINARY; PRT: 537 AA.  
 AC 099988:  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1.  
 OS Porphyra purpurea.  
 OC Mitochondrion.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
 OC NCBI\_TaxID=2787;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Burger G., Saint-Louis D., Gray M.W., Lang B.F.;  
 RT "Complete sequence of the mitochondrial DNA of the red alga, Porphyra purpurea. Inverted repeats, sequence polymorphisms, and cyanobacterial introns."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4 FERROCYTOCHROME C.  
 CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL: AF114794; AAD03116.1; -.  
 DR HSSP: P98002; IAR1.  
 DR InterPro: IPR000883; COX1.  
 DR Pfam: PF00115; COX1.1.  
 DR PRINTS: PR01165; CYCOXIDASEI.  
 DR PROSITE: PS00077; COX1.1.  
 DR Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 DR Respiratory chain; Transmembrane.  
 KW SEQUENCE 537 AA; 59944 MW; EA92FBLCTDFA59DD CRC64;  
 SQ

Query Match 56.9%; Score 1707; DB 8; Length 537;  
 Best Local Similarity 57.2%; Pred. No. 6.3e-86;  
 Matches 310; Conservative 89; Mismatches 101; Indels 42; Gaps 7;

20 RWPSTNHRDGLLYLAAGVGVFSVLTYYMRLEMDPGVQYMCLEGARLIADASOTC 79  
 11 RWPSTNHRDGLLYLAAGVGVFSVLTYYMRLEMDPGVQYMCLEGARLIADASOTC 55  
 80 TANGHLMNVVYTHGLMFEFGVIRALRGFGNYLPLDQAGDAFPRMNNLSFWL--- 136  
 56 LGNHQVYVWVTEHAFLMFEFVMPVVLIGFGNMFVPLMIGADPMPARRNNISFWLPL 115  
 QY 137 ---FLAGTMGVASLEFAPGCGDQLSGVGVNLYPLST--REAGYSMDLAFVHLSGAS 191  
 DB 116 SLCLLGSAM-----VEWAGAGTWLPLPLSSIOSHSGAVDAIFSLHLSGAS 164  
 QY 192 SIMGAINMTTELNMKRAPGMLHKVPLFSMSIFITAMILLALPVLAGATMILLDRNFG 251

```

Db 165 SYLGAINETITFMNRBGOSMYRIPLFVMSILITAFELLLAVLAGALTMILTDNEN 224
QY 252 TTFNPGAGGDPILYXOHLHMFEGHPEVYIIITLPGFIIISHVSTFSKRPVGYLPMYAM 311
Db 225 TTFPDGSGGDPVLXOHLHMFEGHPEVYIIITLPGFIIISHVSTFSKRPVGYIGMIYAM 284
QY 312 VALGVLGVVAHHMYVGSMLTQOSYFMLATVAVPTGIKIFSWATMGGSVEKSP 371
Db 285 LSLGILFIIWAAHMYVGLDVIDRATATMTIINPTGIKIFSWATMGGSVEKSP 344
QY 372 MLMAFGMEFLFTVGVTVGLAAGIDRAVHDYVVAHHEHYVMSLCAIPAIAGIYFYM 431
Db 345 MLFAIGFIFLFTIGLIGLILANSGLDISLHDYVVAHHEHYVMSLCAIPAIAGIYFYM 404
QY 432 PKESGAFPEMAKLFHPTFEGANVTFFPOHPLGROGMPRRYIDYEAFALNNKSSYTG 491
Db 405 EKISGFGYSELGDIHWGTFEGVNLFFEPHHLGLAGMRRIIDYDPSVAGNNIATSYG 464
QY 492 AFLAFASFLFEVTVYLVAGRR-ETRPMPGSEFAD-----TLEWTLSPPAHTEET 544
Db 465 SYVALFSTLFEFLVPTLVTPKRVPRANPW-NEDSKIGSTLLEWISSPFAHYTFNE 523
QY 545 LP 546
Db 524 LP 525

RESULT 10
Q92Y18 PRELIMINARY; PRT; 531 AA.
AC Q92Y18;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Gracilariopsis lemaneiformis.
OG Mitochondrion.
OC Eukaryota; Rhodophyta; Florideophyceae; Gracilariiales; Gracilariaceae;
OC Gracilariopsis.
OC NCBL_TaxID=2782;
RN [1]
RP SEQUENCE FROM N.A.
RA Lang F.B.F., Goff L.J.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
FERRICYTOCHROME C.
CC -1- FACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
EMBL: A118119, AAD17226.1; -.
DR HSSP: P98002; IARI.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
Respiratory chain; Transmembrane.
SQ SEQUENCE 531 AA; 59067 MW; 4C0E66950DE377B6 CRC64;
Query Match 56.9%; Score 1706.5; DB 8; Length 531;
Best local Similarity 57.5%; Pred. No. 6,7e-86;
Matches 308; Conservative 93; Mismatches 106; Indels 29; Gaps 5;

```

```

QY 19 TREMNSTNRKDIGLLVLAAGVGFISVLFYVYMRLEMLDPGVQWCEGARLIADASOT 78
Db 10 SRWIFSTNRKIDISTILLIGAFSGVLCGCMIMRLAQPNQDF----- 55
QY 79 CPVNGSLMMVMYVHGLIMMFPGIPALFGFGNVLPIQIAPAPMAPRNNLSFNLFI 138
Db 56 -LGNHIIYVLLITAHFLMFLFVPMVMTIGFGNNLVPMISSPDAPRLNNISFNLFI 114
QY 139 AGTAMVASLLEPAGDGLSGGVWVLYPPLST--REAGYSMDLAIFVNLSSGSSINGA 196
Db 115 PSLCLLLASAIY-----EVGVGQWTVYRPLSIQSHSGGAVDLAFSLHISGASSILGA 169
QY 197 INMTTFELMMRAGPGLHNVPLFMSISFTAMILLALDVLVLAGATMLTDRNFTTFN 256
Db 170 INFSTILMMRNDGOSMYRMPLEWVSIFETALLLAVVLVLAGATMLTDRNFTTFD 229
QY 257 PAGGCDPLLYXOHLHMFEGHPEVYIIITLPGFIIISHVSTFSKRPVGYLPMYAMVATGV 316
Db 230 PAGGSDPLVXOHLHMFEGHPEVYIIITLPGFIIISHVSTFSKRPVGYIGMIYAMVATGV 289
QY 317 LGRVVAHHMYVGSMLTQOSYFMLATVAVPTGIKIFSWATMGGSVEKSPMLMAF 376
Db 290 LGFTVAAHMYTVGLDVIDRATATMTIINPTGIKIFSWATMGGSVEKSPMLMAF 349
QY 377 GFMELFTVGVTVGLAAGIDRAVHDYVVAHHEHYVMSLCAIPAIAGIYFYMPKPSG 436
Db 350 GFLEFLFTIGLIGLVANSGLDISLHDYVVAHHEHYVMSLCAIPAIAGIYFYMPKPSG 409
QY 437 RAPEWMAKLFHPTFEGANVTFFPOHPLGROGMPRRYIDYEAFALNNKSSYCAPLAF 496
Db 410 VQLEFLGDIHFWSTFEGVNLFFEPHHLGLAGMRRIIDYDPSVAGNNIATSYGAL 469
QY 497 ASFLFETVTVYLVAGRR-ETRPMPGSEFAD-----EADPLEWTLSPPAHTEETLP 546
Db 470 ESTLFEFLVPTLVTPKRVPRANPW-NEDSKIGSTLLEWISSPFAHYTFNE 524

RESULT 11
Q021293 PRELIMINARY; PRT; 531 AA.
AC Q021293;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Reclinomonas americana.
OG Mitochondrion.
OC Eukaryota; core jakobids; Reclinomonas.
OC NCBL_TaxID=48483;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC50394;
RX MEDLINE=97311393; PubMed=9168110;
RA Lang B.F., Burger G., O'Kelly C.J., Cedergren R., Golding G.B.,
Lemieux C., Sankoff D., Tumeit M., Gray M.W.;
RT "An ancestral mitochondrial DNA resembling a eubacterial genome in
miniature."
RL Nature 387:493-497(1997).
DR [12]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50394;
RA Lang B.F., Burger G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4

```







Db 227 YDPAGGSRPVLVYQHILFMFEFGHPIREYVILILPAFGILISQVVAFSRKPVPFGYLCGMVALLSI 286

Oy 315 GVLGIVVVAHHMYTVGMSLTQOSTFMALMTVAVPVGIKTFSKATMMGSGVERKSPMLM 374

Db 287 ILLGLGLVAAHHMFYGVMDVTRAFVTAIIIAVPGIKVFSWATAMMGSLHLOTPLMF 346

Oy 375 AFGFMFLFTVGVGVGIVLAQGLDRAVHDPYVVVAHFHVMYSIGATFALFPAGIYPMKFE 434

Db 347 ALGFLILFTVGGGLGVGIVLSNGLDVAHDHYVVAHFHVLMSGLGFLGFAFYIWKIKI 406

Oy 435 SCRAREPMAALHFWTFEFGANVTFFPQHFVLRGQMPRRYIDYPEAFALMKNVSSGAFI 494

Db 407 SKQKPEVLAQGHFLITFLTGLVNLFFEFMHFLGLAGIPRRIPDYPAVAGMNLVATGSGYV 466

Oy 495 A-FAFLEFPIFYVTVLVAGRRFRPNQGEFAD-----TLEWLPSPRHHTET 544

Db 467 TLFGLLFFLYIVY-TLLEG-QVCEPDNPVATLDDQTSNGSTTLEWLVSPPAYHTIVE 524

Oy 545 LP 546

Db 525 LP 526

RESULT 14

P92800 PRELIMINARY; PRT; 528 AA.

AC P92800:

DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).

OS COXI.

GN Pyraliella littoralis.

OC Mitochondrion.

OC Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;

OC Ochlosporaceae; Pyraliella.

NCBI\_TaxID=2885;

RN [1]

RP SEQUENCE FROM N.A.

RA Fontaine J., Goux D., Kioareg B., Loiseaux-de Goe S.;

RL J. Mol. Evol. 0:0-0(0).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: Z72500; CAA96581.1. -

DR HSSP: P98002; IARI.

DR Mendel: 21583; Pylii:coxi;21583.

DR InterPro: IPR002106; AA\_TRNA\_Ligase\_II.

DR InterPro: IPR000883; COXI.

DR Pfam: PF00115; COXI; 1.

DR PRINTS: PR01165; CYCOXIASE1.

DR PROSITE: PS00339; AA\_TRNA\_Ligase\_II\_2; UNKNOWN\_1.

DR PROSITE: PS00077; COXI; 1.

DR COPPER; Heme; Inner membrane; Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane.

KW SEQUENCE 528 AA; 58015 MW; 11E3DDBD9A055BAE CRC64;

SW

Query Match 55.3%; Score 1659.5; DB 8; Length 528;

Best Local Similarity 57.5%; Pred. No. 2,4e-83;

Matches 308; Conservative 85; Mismatches 106; Indels 37; Gaps 8;

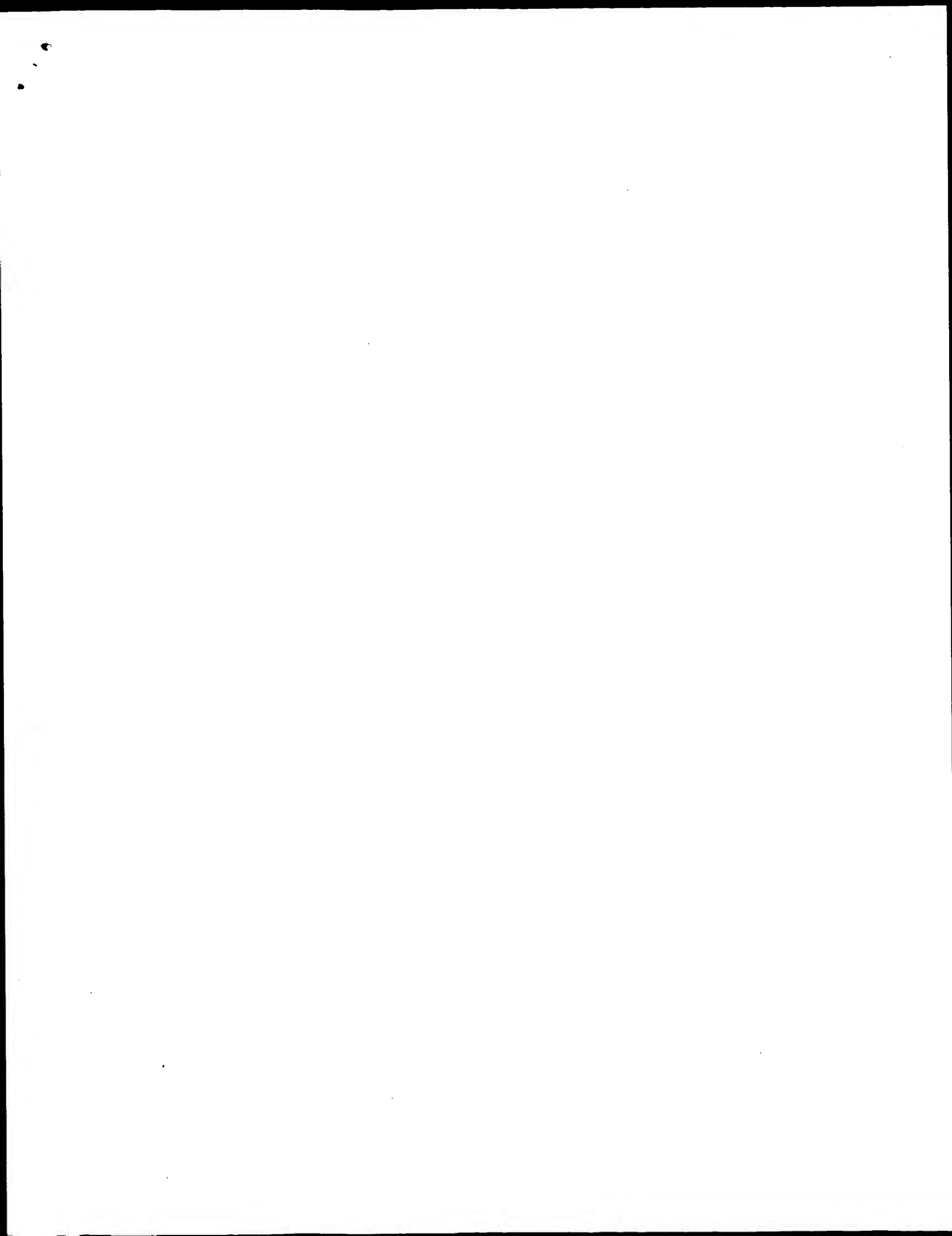


Thu Dec 27 08:21:36 2001

us-09-712-768-2.rpt

---

Page 11



Thu Dec 27 08:21:15 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 08:09:34 ; Search time 2710.12 Seconds  
(without alignments)  
10190.057 Million cell updates/sec

Title: US-09-712-768-1  
1674  
Sequence: 1 atgcagacgcccgcactca.....acaagcatccctgcactaa 1674

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: gb-ba:\*  
2: gb-hcg:\*  
3: gb-in:\*  
4: gb-om:\*  
5: gb-ov:\*  
6: gb-pat:\*  
7: gb-ph:\*  
8: gb-pl:\*  
9: gb-pr:\*  
10: gb-ro:\*  
11: gb-sts:\*  
12: gb-sy:\*  
13: gb-un:\*  
14: gb-vi:\*  
15: em-ba:\*  
16: em-fun:\*  
17: em-hum:\*  
18: em-in:\*  
19: em-om:\*  
20: em-or:\*  
21: em-ov:\*  
22: em-pat:\*  
23: em-ph:\*  
24: em-pl:\*  
25: em-to:\*  
26: em-sts:\*  
27: em-sy:\*  
28: em-un:\*  
29: em-vi:\*  
30: em-htgo-hum:\*  
31: em-htgo-inv:\*  
32: em-htgo-rod:\*  
33: em-htg-hum:\*  
34: em-htg-inv:\*  
35: em-htg-rod:\*  
36: em-htg-other:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1674	100.0	1674	6	AX113986	AX113986 Sequence
2	1091.4	65.2	2629	1	PDCYCA	Y07533 P.dentifri
3	991	59.2	3211	1	PDCOX2	X05829 P.dentifri
4	979.2	58.5	2166	1	RSCOXSI	B2645 R.sphaeroid
5	754.4	45.1	286550	1	SME591785	AL591785 Sinothlzo
6	750.2	44.8	11188	1	AE006001	AE006001 Caulobact
7	740.4	44.2	10835	1	AE008010	AE008010 Agrobacte
8	719.6	43.0	346510	1	AP003011	AP003011 Mesorhizo
9	709.2	42.4	4380	1	NMCOXABC	X8566 B.wingradis
10	708.8	42.3	1830	1	BJUCOXAG	X54318 B.japonicum
11	708.8	42.3	2000	1	BJUCOXA	X54800 B.japonicum
12	708.8	42.3	8121	1	BJA242592	AJ242592 Bradyrhiz
13	708.8	42.3	31495	1	BJU33883	U33883 Bradyrhizob
14	654	39.1	1707	1	RLCOXA	X74341 R.leguminos
15	515.6	30.8	12829	1	AE004449	AE004449 Pseudomon
16	456.6	27.3	1406	8	AF020565	AF020565 Ginkgo bl
17	454.4	27.1	1412	8	AF020564	AF020564 Ephemera v
18	450.6	26.9	1415	8	AY009436	AY009436 Calycanth
19	450.6	26.9	4815	8	MIRSCOXI	X57692 R.sativus m
20	449	26.8	1413	8	AF193955	AF193955 Calycanth
21	449	26.8	166924	8	MATGENB	Y08502 A.thaliana
22	447.6	26.7	1415	8	AF020585	AF020585 Ephemera d
23	447.4	26.7	1584	8	PTU77623	U77623 Populus tre
24	445.8	26.6	1413	8	AF193962	AF193962 Buxus sp.
25	444.2	26.5	1415	8	AY009437	AY009437 Canella w
26	444.2	26.5	1415	8	AY009447	AY009447 Nelumbo l
27	442.6	26.4	1409	8	AF020566	AF020566 Gnetum le
28	442.4	26.4	1409	8	AF020576	AF020576 Sciadopit
29	441.4	26.4	1881	8	MTRACOI	Y00417 Wheat mitoc
30	441.4	26.4	2048	8	ACUA6764	U46764 Aegilops co
31	441.4	26.4	3059	8	MTRACOXI	X56186 Triticum el
32	441	26.3	1409	8	AF020572	AF020572 Ophioglos
33	440.8	26.3	1415	8	AF020584	AF020584 Melwitsch
34	440	26.3	1701	8	BVCOXI	X57693 B.vulgaris
35	440	26.3	169799	8	AP000396S2	AP000397 Beta vulg
36	439.8	26.3	2149	8	MUCOXI	X02660 Maize mitoc
37	439.8	26.3	2223	8	M14454	M14454 Sorghum bic
38	439.8	26.3	2545	8	SRGWTCOXI	M14453 Sorghum bic
39	439.6	26.3	1415	8	AF020573	AF020573 Phyllocla
40	436.8	26.1	2481	8	AB024985	AB024985 Beta vulg
41	436.6	26.1	1955	8	MISCOXI	X15990 Rice mitoch
42	434.6	26.0	1413	8	AF193968	AF193968 Geranium
43	434.6	26.0	16537	5	AB047553	AB047553 Plecoglos
44	433.6	25.9	1413	8	AF193967	AF193967 Hypseochn
45	428.2	25.6	1415	8	AY009454	AY009454 Aritsaema

## ALIGNMENTS

RESULT	1					
AX113986	AX113986	1674 bp	DNA			
LOCUS	Sequence 1 from Patent EP1103603.					
DEFINITION	AX113986					
ACCESSION	AX113986					
VERSION	AX113986.1	GI:13940146				
KEYWORDS						
SOURCE						
ORGANISM						
	Gluconobacter oxydans.					
	Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;					
	Gluconobacter.					
REFERENCE	1 (bases 1 to 1674)					
AUTHORS	Asakura,A., Hoshino,T. and Shinjoh,M.					
TITLE	Cytochrome c oxidase complex from gluconobacter oxydans					
JOURNAL	Patent: EP 1103603-A 1 30-MAY-2001;					
	F. HOFFMANN-LA ROCHE AG (CH)					
FEATURES						
SOURCE	location/Qualifiers					
	1..1674					
	/organism="Gluconobacter oxydans"					
	/db_xref="taxon:442"					

CDS 1. 1674  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37869.1"  
/db\_xref="GI:13940147"  
/translation="MADAAIHGHHEKQFPTRMFSTNKHIDILYLIVAGVVGFI  
SVLFTVMRLMDPGVOYMCLEGRRLIADASQCTANGHLVAVTVYITLMFPG  
IPALFGGNTYMLPIOTIGAPDAPRPMNNTSMPLTACGACVASTFAPGGDGLSG  
VGMVLPPLSTREASTSMDLAFVAVHLSGASIMATIMTTFELMRAPGHTLHKVPL  
FWSMIFITAMLLLPVLAGITMLLTDNRNGTFENPAGGDEILYOLHMFGRPL  
EYVITILPFGFISHVSTFESKRPVGYLPMYAVALGVGLVMAHMTVMGSLT  
OOSYMLATMVAIVPTGIKIFSMITMNGSGVEFPMLEFMEFLTGVGVIVL  
AAGLDRAYHDYVVAHRYHVMISGLAIFAIPTAGIYFPMPEFSGAPREMAKLFHMT  
FRIGANVFEPHFLGROGMPPRYIDYDEAFALMKVSSYSGAFIAPAFEFYIVFY  
TLVAGRRETRPMPMEPRADTLEMTLPSPRPATTFETLKRSDMKRPSH"

BASE COUNT 270 a 493 c 468 g 443 t

ORIGIN

Query Match 100.0%; Score 1674; DB 6; Length 1674;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgagcagacccgcacatcaagcgcaatgacacacatgaagaagcctcttcacgacg 60  
1 ATGGCAGACCCGCCCATTCACGGCCATGACACCATGAGACGACGCTTTCACGCCG 60

QY 61 tggatcatgctgac 120  
61 TGGTTCATGCTGACACCAACCAAGACATCGCTCTATACCTTTACGCGCTGCTGTT 120

QY 121 gttggtttacatcctgctcgttcacacacacacacacacacacacacacacacacac 180  
121 GTTGTTTACATTCCTCCTCCTGTTACCGCTTACATCGCGCTTACGTTGATGATCGG 180

QY 181 gttcagatcatcgtcgttcacacacacacacacacacacacacacacacacacacac 240  
181 GTTCAGTACATGCTGCTTGAAGCGCACGCTGATCGCGGATCCTCGCAGACATGTAGC 240

QY 241 ggcgaacggaac 300  
241 GCGAAGCGAACCTGTGGAACGTCATGCTTACCTACCATGCTATTTGATGATGCTTCTT 300

QY 301 gttggtttacatcctgctcgttcacacacacacacacacacacacacacacacacac 360  
301 GTTGTTTACATTCCTCCTCCTGTTACCGCTTACATCGCGCTTACGTTGATGATCGG 360

QY 361 gttcagatcatcgtcgttcacacacacacacacacacacacacacacacacacacac 420  
361 GTCGCGATATGCGCTTCCCGCTATGAAACATGCTGCTGCTGCTGCTGCTGCTGCTG 420

QY 421 accgagatgagcgttcgttcacacacacacacacacacacacacacacacacacacac 480  
421 ACCGCGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

QY 481 gttggtttacatcctgctcgttcacacacacacacacacacacacacacacacacac 540  
481 GTTGTTTACATTCCTCCTCCTGTTACCGCTTACATCGCGCTTACGTTGATGATCGG 540

QY 541 ggcgaacggaac 600  
541 GCGAAGCGAACCTGTGGAACGTCATGCTTACCTACCATGCTATTTGATGATGCTTCTT 600

QY 601 agcagacatcctggaac 660  
601 AGCAGACCTCTTGAACACATGCGCGCCGCGATACCGTGCACAAAGCGCGTTGCTCTCG 660

QY 661 tggatcatgctgac 720  
661 TGGTTCATGCTGACACCAACCAAGACATCGCTCTATACCTTTACGCGCTGCTGTT 720

QY 721 atcaacatgctgac 780  
721 ATCAACATGCTGACACCAACCAACATCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

Db 721 ATCAACATGCTGACACCAACCAACATCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

QY 781 ggtgacccgcatcgtgac 840  
781 GGTGACCCGATTCGTATACCAACATCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 840

QY 841 atcaacatgctgac 900  
841 ATCAACATGCTGACACCAACCAACATCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

QY 901 gttcagatcctgac 960  
901 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

QY 961 gttcagatcctgac 1020  
961 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020

QY 1021 ctggcagac 1080  
1021 CTGGCAGACACATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080

QY 1081 atgtaggacgac 1140  
1081 ATGTGGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140

QY 1141 ctgttccac 1200  
1141 CTGTTCACCGTGGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

QY 1201 taltacagac 1260  
1201 TATACGACACACTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260

QY 1261 ttgtagac 1320  
1261 TTGCGGATCTTCCGCGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320

QY 1321 gaattgac 1380  
1321 GAATGGGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380

QY 1381 ccgcagac 1440  
1381 CCGCAGACACTTCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440

QY 1441 ttcgtagac 1500  
1441 TTCGCGCTGTGGAACAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500

QY 1501 tttcctcatcgtgac 1560  
1501 TTTCTTCAATGCTGATTTTGTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

QY 1561 ccgttggagcgaatctgcgcgaatcgtggaatcgtggaatcgtggaatcgtggaatcgtg 1620  
1561 CCGTGGGCGGGAATTCGCGATACGCTGGAATGAGACGCTGCGCATACCGCCCTCCG 1620

QY 1621 acgttgcgaac 1674  
1621 ACGTTTCGAACACCTTCGCCAAGCGCTGCGAGCTGGGCAAGCAATCCCTCGCACTAA 1674

RESULT 2

PCYCXA 2629 bp DNA BCT 23-JAN-1992

DEFINITION P.denitrificans cyc A and cta DII genes for cytochrome c 550 and

cytochrome c oxidase beta subunit I.

ACCESSION Y07533

VERSION Y07533.1 GI:45477

KEYWORDS cta DII gene; cycA gene; cytochrome c oxidase; cytochrome c-550.

SOURCE Paracoccus denitrificans.

## ORGANISM

Paracoccus denitrificans  
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;

## REFERENCE

1 (bases 1 to 2629)

## AUTHORS

Raitio, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (17-APR-1990) M. Raitio, DEPT OF MEDICAL CHEMISTRY,  
UNIVERSITY OF HELSINKI, SILTAVUORENENGERR 10, 00170 HELSINKI 17,  
FINLAND

## REFERENCE

2 (bases 1 to 2629)

## AUTHORS

Raitio, M., Pispä, J. M., Metsä, T. and Saraste, M.

## TITLE

Are there isoenzymes of cytochrome c oxidase in Paracoccus

## JOURNAL

denitrificans?

## MEDLINE

FEBS Lett. 261 (2), 431-435 (1990)

## FEATURES

90184495

## SOURCE

1. 2629  
Location/Qualifiers

## FEATURES

/organism="Paracoccus denitrificans"

## FEATURES

/strain="PD122"

## FEATURES

/sub\_strain="ssp. MR-3"

## FEATURES

/db\_xref="taxon:266"

## FEATURES

/clone\_lib="EMBL3 lambda"

## FEATURES

233..2544

## FEATURES

/gene="cyc A"

## FEATURES

233..292

## FEATURES

/gene="cyc A"

## FEATURES

233..700

## FEATURES

/gene="cyc A"

## FEATURES

/EC\_number="1.9.3.1"

## FEATURES

/note="cytochrome-c oxidase"

## FEATURES

/codon\_start=1

## FEATURES

/transl\_table=1

## FEATURES

/product="cytochrome c 550"

## FEATURES

/protein\_id="CA68820.1"

## FEATURES

/db\_xref="GI:45478"

## FEATURES

/db\_xref="SWISS-PROT:P00096"

## FEATURES

/translation="MKISTYATLAITLALPAAQSDAAKGEKEFKKCKACHMIOAP  
DGDIDIKGKTPNLYGVGRKIASBEGFYGIEIVAKNPDLTETADLLEIYVD  
PKPMLVMTDCKAKTKMFEKMKGNADVAVFLAQNSPDAGGDEMAAEESEN"

## FEATURES

293..2544

## FEATURES

/gene="cyc A"

## FEATURES

/EC\_number="1.9.3.1"

## FEATURES

/note="cytochrome c 550"

## FEATURES

/product="cytochrome-c oxidase"

## FEATURES

731..765

## FEATURES

/gene="cyc A"

## FEATURES

871..2547

## FEATURES

/gene="cta DII"

## FEATURES

871..2547

## FEATURES

/gene="cta DII"

## FEATURES

/EC\_number="1.9.3.1"

## FEATURES

/note="cytochrome-c oxidase"

## FEATURES

/codon\_start=1

## FEATURES

/transl\_table=1

## FEATURES

/product="cytochrome c oxidase beta subunit I"

## FEATURES

/protein\_id="CA68821.1"

## FEATURES

/db\_xref="GI:45479"

## FEATURES

/db\_xref="SWISS-PROT:P98002"

## FEATURES

/translation="MADAHVGHGDDHDTGPFTRMFMSTNHNKDIGILVFTAGIVGL  
ISVCFYVMMLOHPOVOYMLLEGATLIDASECTNGLMNWITTYGVMMFV  
VITALRGKNTFRPLHIGAPDPAFRPLNLSITWYVCGVALVASLLAGGDDOKS  
GKAVNLPPLSTLEAGISMDLAFVAVHSGASSILGAINITFFLNKRAFGMTLFV  
LFMSVFTITAMLLSLPLVLAGITMLMDRNGTOFFDPAGGDDVLYOHLMPFGH  
PEVYIILPGFGIISHVSTFARKPIFGYLPRLAMAALGILFVVAMHMYTAGVSL  
TOOAYFLATMTIAVPTGIVKFSMATMGSGIEFKIPMLMAFGPLFLFVGGVTV  
LSQAPLDVYHDTYVYHMYHVSAGVGTGAGVYWTGKSGSGOYPRMAOGLFM  
MFTGSLIFEPHILGROGMPRIYIDYPERFYNNISSIGATISFASLFTGTIVF  
YTLPAKRNVVPMNMDHDTLEMTLPSPPEHTFTLKRBDWDRHAH"

## FEATURES

2386..2615

## FEATURES

872 c 771 g 529 t

## FEATURES

terminator

## FEATURES

BASE COUNT 457 a

## FEATURES

ORIGIN

Query Match 65.2%; Score 1091.4; DB 1: Length 2629;  
Best Local Similarity 78.6%; Pred. No. 9.1e-219;  
Matches 1318; Conservative 0; Mismatches 356; Indels 3; Gaps 1;

QY 1 atggcaagacgcccgcattcagcgcca---tgaccacatgagaagaacgcttctcag 57  
Db 871 ATGCAGACGCGAGCGGTTCAACGGCCAGGAGGACATCATGACACCGCGGCTTCCAC 930  
QY 58 cgcgtggtcattgctaccacaacaaagacatgctgctataccttctgagcgctgct 117  
Db 931 CGGTGCTTCAATGTCACAAACACACAAAGATATGCTTATCCTTACGTCAGCGCGG 990  
QY 118 gttgtggttcaattccgctccgctgctacacgcttcaagcttcaagcttcaagctt 177  
Db 991 ATGCTGCGGCTGATCTCGGATGCTTACCGCTCTATATGCGGATGACACTGACATCC 1050  
QY 178 gttgttgaatgacatgctgcttgaagagcgcaagcttgaagcttgaagcttgaag 237  
Db 1051 GGCCTGCAATATCATGCTCCGGAAGCGCGGCTCTATCCGACGCGCTCGCGGATGC 1110  
QY 238 acgcaagacgagacacgctgctgagcgtatgcttaccacacatgcttctgagcttc 297  
Db 1111 ACCCGAAGCGACACCTGCTGGAACGTATGATATCATACCGCGGCTGATGATGTC 1170  
QY 298 ttgtggtatcccgcatgttctggtgttctgtaacatctgtagcgcgtgcaatc 357  
Db 1171 TTGCTGCTGATCCCGGACGCTTGGGGGTTTGGGACACTATTTCATGCCGCTGATTC 1230  
QY 358 ggcgtccgagatagcctcccgcgatagaacaacctgctgctgctgcttcaatgctc 417  
Db 1231 GGCCTCCCGACATGCGCTTCCCGGCTGACACACTCTCTACTAGATGATGCTGTCG 1290  
QY 418 ggtacgagctgagcggtgctgctgctgctgctgctgctgctgctgctgctgctg 477  
Db 1291 GGCCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1350  
QY 478 ggcgttctggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 537  
Db 1351 GCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1410  
QY 538 ctgcgattctcgcggttcaatctgctgctgctgctgctgctgctgctgctgctgct 597  
Db 1411 CTGCGATCTTTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1470  
QY 598 atcagcagcttctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 657  
Db 1471 ATCACACCTTCTCTCAACATGCGCGCACCGGATGACGCTGTTCAGAGTCCGCTG 1530  
QY 658 tctggtcgaactcttcaacagcgttgcgtatcctgctgctgctgctgctgctgctg 717  
Db 1531 GCGTGTGCGCTTTCATCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1590  
QY 718 gcaatcacatgctgctgagcagcagcagcagcagcagcagcagcagcagcagcag 777  
Db 1591 GCGATCACCATGCTGCTGATGAGCACCGCAACCTTGGCAGCAGTTCATCCGCG 1650  
QY 778 ggcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 837  
Db 1651 GCGCGGACCGCGGCTGCTTACACGACATCTGCTGCTTCTGCGCATCCAGAGCTCT 1710  
QY 838 atcatcattctgcgcgcttctgcatcatalcagccatgctgctgcttctgcaaaag 897  
Db 1711 ATCATCATCTGCGCGGCTTGGCATCATCAGCAGCATGCTGCGCTTGGCAGAG 1770  
QY 898 cgcgtctggttacttctgagcagcagcagcagcagcagcagcagcagcagcagcag 957  
Db 1771 CCGATCTTCCGCTACCTGCGGATGCTGCGGATGCGGCGGCGGCGGCGGCTTCC 1830  
QY 958 gtcgtctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1017  
Db 1831 GTCGTCTGGCGGACGACATGTCACGCGCGGATGCTGCTGACCGACGAGGCTTTC 1890  
QY 1018 atgctgccaacatgctgagcagcagcagcagcagcagcagcagcagcagcagcag 1077





QY 191 tgcgcttgaagcgacgctctgacgagatgcccgcagacatgtaacggaagcagac 250  
 DB 961 tctgcttgaagcgacgctctgacgagatgcccgcagacatgtaacggaagcagac 1020  
 QY 251 accgtgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 310  
 DB 1021 accgtgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080  
 QY 311 cggagcttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 370  
 DB 1081 cggagcttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1140  
 QY 371 tggccttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 430  
 DB 1141 tggccttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
 QY 431 ggcgtgcttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 490  
 DB 1201 ggcgtgcttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260  
 QY 491 tctgtacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 550  
 DB 1261 tctgtacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320  
 QY 551 cggctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 610  
 DB 1321 cggctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380  
 QY 611 tgaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 670  
 DB 1381 tgaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440  
 QY 671 ttaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 730  
 DB 1441 ttaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
 QY 731 tgcctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 790  
 DB 1501 tgcctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560  
 QY 791 tctgtacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 850  
 DB 1561 tctgtacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1620  
 QY 851 cggctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 910  
 DB 1621 cggctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1680  
 QY 911 accgtgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 970  
 DB 1681 accgtgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740  
 QY 971 accgtgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1030  
 DB 1741 accgtgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1800  
 QY 1031 tgggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1090  
 DB 1801 tgggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1860  
 QY 1091 gctcgttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1150  
 DB 1861 gctcgttgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1914  
 QY 1151 tgggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1210  
 DB 1915 tgggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1974  
 QY 1211 cctatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1270  
 DB 1975 cctatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2034

QY 1271 tgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1330  
 DB 2035 tgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2094  
 QY 1331 caaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1390  
 DB 2095 caaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2154  
 QY 1391 tctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1450  
 DB 2155 tctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2214  
 QY 1451 ggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1510  
 DB 2215 ggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2274  
 QY 1511 tgaacttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1570  
 DB 2275 tgaacttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2334  
 QY 1571 aatcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1630  
 DB 2335 aatcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2394  
 QY 1631 cgcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1694  
 DB 2395 cgcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2458

RESULT 4  
 LOCUS RSCOX1 2166 bp DNA BCT 09-APR-1992

DEFINITION R.sphaeroides gene for cytochrome oxidase subunit 1.  
 VERSION X62645.1 GI:46417  
 KEYWORDS cytochrome oxidase subunit 1.  
 SOURCE Rhodobacter sphaeroides.  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.

REFERENCE 1 (bases 1 to 2166)  
 AUTHORS Shapleigh,J.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-OCT-1991) J.P. Shapleigh, University of Illinois, Dept. of Biochemistry, 505 S. Mathews, Urbana IL 61801, USA  
 REFERENCE 2 (bases 1 to 2166)  
 AUTHORS Shapleigh,J.P. and Gemmis,R.B.  
 TITLE Cloning, sequencing and deletion from the chromosome of the gene encoding subunit I of the a<sub>33</sub>-type cytochrome c oxidase of Rhodobacter sphaeroides  
 JOURNAL Mol. Microbiol. 6 (5), 635-642 (1992)

JOURNAL 92204019  
 MEDLINE  
 FEATURES  
 Source location/Qualifiers

CDS  
 1..2166 location="Rhodobacter sphaeroides"  
 /organism="Rhodobacter sphaeroides"  
 /strain="GA"  
 /db\_xref="taxon:1063"  
 207..1904  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome oxidase subunit 1"  
 /protein\_id="CAA44514.1"  
 /db\_xref="GI:46418"  
 /db\_xref="SWISS-PROT:P33517"

/translation="MADAAIRHEDHRRFETRMWSTNRHNDIGVLYLFTGGLVGLIS  
 VAFIVRMELNAPGVQEWCAHLESGVKGFFQSLWPSAVENCTPNGHLMWMTGH  
 GILMFVVPALPFGFNGYFMPHLHGAIDNAEPKNNLSVLYVAGSLAVASLAP  
 GGNGQSGSIGWVLYPLSTSESGVSTDAIFAVHLSGASLILGAINMTTFLNLRAP  
 GMTMRKVPLEAFMSIEVTAMLLALPVLALPDLTLDSEFTFEPSGGSDPYLYO  
 HLMRGHEVYIVIPARGIVSHVATPAKPDITGILPMYAMATVLCVGFVVAHH  
 MTAGSLTQOSTFPMATNVAIVPPIGKIFSLIATIMWGSIELKPMIALGFLPLT  
 VGVGVTVISQASVDRYVDHYVVAHVPVSLDAVPSIFAGSISGIGKMSGRQYPE  
 WGKLFHMFVFGANLTFPPHFLGRCMPRRYIDYPEAFATWNVSSILGFLSPASF



PUBMED 11481430  
 REFERENCE 2 (bases 1 to 286550)  
 AUTHORS Gouzy J.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELLIO EU Consortium  
 COMMENT

States of America. 98 (17), 9877-9882 (2001)  
 11481430  
 2 (bases 1 to 286550)  
 Gouzy J.  
 Direct Submission  
 Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELLIO EU Consortium  
 COMMENT  
 Laboratoire de Biologie Moléculaire des Relations  
 Plantes-Microorganismes, UMRI215-CNRS-INRA, BP27, F-31326 Castanet,  
 France, Laboratoire de Génétique et Développement UMRI061-CNRS,  
 Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,  
 France, GATC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany,  
 Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,  
 D-33615 Bielefeld, Germany, Unité de Biochimie physiologique 25,  
 Université catholique de Louvain, Place Croix du Sud 2, Bte 20,  
 B-1348 Louvain-la-Neuve, Belgium, Unité de Microbiologie, Faculté  
 des Sciences Agronomiques de Gembloux, Avenue Mareschal Juin 6,  
 B-5030 Gembloux, Belgium. E-mail: Jérôme.Gouzy@toulouse.inra.fr  
 http://sequence.toulouse.inra.fr/mellioti.html.

## FEATURES

source

Location/Qualifiers

1..286550

/organism="Sinorhizobium meliloti"

/strain="1021"

/db\_xref="taxon:382"

70..813

/gene="gstI1 OR SMC00916"

70..813

/BC-number="2.5.1.18"

/function="miscellaneous"

/note="Product confidence : putative"

Gene name confidence : putative

Predicted by Codon usage

Predicted by Homology

Predicted by Framed

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"

/protein\_id="CAC45361.1"

/db\_xref="GI:15073720"

/translation="MADLSFPTARKPKNPDIIDLYSPFNGKVKVSTALEELGIA  
 YEHRISFDNEKSPFELSINPGRIPALIDNGDPKPIGLFESGAILIYAEKGTG  
 KLPADARARYETLCWVFQMGIGPMFGQGHFPEADKANNPYPVORYDEAKR  
 LGLVEARLEGRQMLMGKRNTPSPSPSPGKRGSSSVAVPRCSTLRASRMSTGMI  
 AAMHALPRRAWSRRASRSRSLPAGD"

588..779

/gene="SMC00915"

588..779

/gene="SMC00915"

/function="miscellaneous: hypothetical/partial homology"

/note="Product confidence : hypothetical"

Gene name confidence : hypothetical

Predicted by Codon usage

Predicted by Homology

Predicted by Framed

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="HYPOTHETICAL PROTEIN"

/protein\_id="CAC45362.1"

/db\_xref="GI:15073721"

/translation="MADGGEETIADIATYVWEGARKFYGAENVLDYKSPNMDV  
 DRGLRPAAGKGMPIRKE"

867..2084

/gene="SMC00914"

867..2084

/gene="SMC00914"

/function="small molecule metabolism"

/note="Product confidence : putative"

Gene name confidence : hypothetical

Predicted by Codon usage

## repeat\_region

gene

CDS

predicted by Homology  
 Predicted by Framed  
 /codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="PUTATIVE OXIDOREDUCTASE PROTEIN"  
 /protein\_id="CAC45363.1"  
 /db\_xref="GI:15073722"

/translation="MSGRLVVGCGQAFALVAKLRALKDMRPITVVAEASLPYORP  
 PLSKYLLREMLDLRLYRPEAVYAEHEIDRLSTTVYRDLARQVALSDSMLYE  
 TLAFATGAPRRRLPAVGGDLVAGVYRDPDRADRLAEEMGPRVVLVGGYIGLEA  
 AAVAFSGEAGTVYEMADRIILQRYSAATSAIVREIHRSHGVDIRRIGHLIGDNG  
 RYAAELSGSVYIPDVIYVIGVANAADLAHEAGIETANGIYVDSHSGTSPTTYAM  
 GGCADVPMGKRIRESVYMAQDAEVAVAIVLAGITDPRDPKPRMSQYDVKQDIAG  
 FGIHDETVLRGQGRGVSVMYFRQGLIADVAINDAKAVATYTKLLKAGATPDRLL  
 LADPYDIALALV"

complement(2114..2572)

/note="Sm-2 OR SMC04876"

REPEAT SM-2

/evidence=not\_experimental

complement(2585..4222)

/gene="groEL1 OR SMC00913"

complement(2585..4222)

/gene="groEL1 OR SMC00913"

/function="cell processes: chaperoning"

/note="Predicted by Homology"

Predicted by Homology

Predicted by Framed

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="60 KD CHAPERONIN A PROTEIN"

/protein\_id="CAC45364.1"

/db\_xref="GI:15073723"

/translation="MAAKEVFRGRSAREKMLRGVDILADAVKVTLGPKGRNVVDKSP  
 GAPRITKDGVSVAKEIELEDPENGAQKRVKASTNDIADGTTTATVLAQVIRE  
 GAKAAVAGNPMDLKRGDILAAEVYKDLAKKINISDEAVAGGTISAVGKQIGL  
 DIAFMOKVNGSVITVEBAKTAETELVEVEMQDRGLTSPYPTNPERKVADEDA  
 FLIHRKKSINQAMLPITAEVYQTKPELLITAEVEBEALATLVNKLKGLKIAAV  
 KAPGFGRRKAMLEIDITLTGTVTSEDDIKLESTVLDMLGRARVITKENTIYD  
 GAGQKSDIEGRAVQIKQIETTSYDEKLOERLAKLAGGAAVTRVGATVEVEKER  
 KRIDDALNATRAAVQEGIVPGGVALLRSSVKTIVKGEENDODAGVAVIRRALQSPA  
 RQIVENAGSEASIVVGIILEKNTDDPGVNAQGEYGDVMIANGIIDPVVVRTALQDAA  
 SVASILLITTEAMIAELPKPAMRGMGCMGCMGM"

complement(4298..4594)

/gene="groES1 OR SMC00912"

complement(4298..4594)

/gene="groES1 OR SMC00912"

/function="cell processes: chaperoning"

/note="Predicted by Codon usage"

Predicted by Homology

Predicted by Framed

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="10 KD CHAPERONIN A PROTEIN"

/protein\_id="CAC45365.1"

/db\_xref="GI:15073724"

/translation="MASTNFRPLHDVVVVRVSESEKTKGIIIPDTAKERPQGEIV  
 AVSGSARDSCGVPLDVAKDRIILFCWMSGTEVKINEDLLIMKEMDIMGVIG"

4863..5690

/gene="SMC00911"

4863..5690

/gene="SMC00911"

/function="miscellaneous: hypothetical/global homology"

/note="Product confidence : hypothetical"

Gene name confidence : hypothetical

Predicted by Codon usage

Predicted by Framed

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

gene  
 CDS

gene  
 CDS

gene  
 CDS

gene  
 CDS

gene  
 CDS

gene  
 CDS





```

gene      complement(6601..7098)
CDS       /gene="CC3412"
          complement(6601..7098)
          /gene="CC3412"
          /note="identified by match to protein family HMM"
          /codon_start=1
          /transl_table=11
          /product="acetyltransferase, GNAI family"
          /protein_id="AAK25374.1"
          /db_xref="GI:13425124"
          /translation="MSRLSLPLADLADNAPGALLALINNAHVELSMLEPERLAHLV
          AQAFLARRVGVADALLTFEDHDADSPNLFMRFRPPFVYVDRVVVADARSGLA
          RALYDLDLPAKAAGHARIVCEVNSDPNPVSDAFHALLGFEVPVGTALIHGKKTVY
          LERRL"
gene      complement(7386..9266)
CDS       /gene="CC3413"
          complement(7386..9266)
          /gene="CC3413"
          /note="identified by match to protein family HMM"
          /codon_start=1
          /transl_table=11
          /product="peptide transporter"
          /protein_id="AAK25375.1"
          /db_xref="GI:13425125"
          /translation="MNIIVAGILVLTVTGIPVLQLKHPRLIICELAEEMERES
          YGMRALILFYLTHQLFDPKASGHSYALVYVPLIGFLADKMLGTRKAVAFG
          AVLVAHLAMAEIEGKPAVQTLTYGATVEQAEGRGKLVDDKAEVATKAVF
          GFETKGLPANAPLPAVLPAKADYKLDVGRDPLINIFSLIILIGVGLKPNIST
          IYGLXKPGDPRDRPGLTYXGILGSPMASIICGLIGVVMWAGGGLAGVGMALG
          FIVFVGLKRWLDKAEPEPKYKLEKVAQPIRELALITGLSLVGLVGFPIYORALY
          GVAINIGLASLGYIGFEMFTKAEERERLAVLIFGAVENTLFEQGSLSLEF
          AATNVDLVRPRLFSGAVTLATPDOLAKAGIDAFTFVNDSENFQOALNAGW
          ILIFAPVAWVTFGLARGKDPNPVTKFGLALIQVAGFLLDNGSTPDAQAFKPLI
          FLVLMVHTTGEICMSPVLSOMTKLSPASVSPMAWVFLAVAIAGWVGSKLAGLA
          ATEYGGQVLDGALALASIKFNITIGLVAVALIGFLLSPVLKKNHGDNDVGNAPK
          KA"
gene      complement(9726..11087)
CDS       /gene="CC3414"
          complement(9726..11087)
          /gene="CC3414"
          /note="similar to GP:3688156; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="glutamate-cysteine ligase"
          /protein_id="AAK25376.1"
          /db_xref="GI:13425126"
          /translation="MADTRASVOEROLITLDDLTAYFAKSGKPEARVGAHEHKKGFYL
          GSHEVPYEGDKGVALLGLGDRFGMTPEMESPTTIGLERNCANVSLRPGQFELSGA
          PLSTMEICEETGQHLDEVKTVADDELGLGFTGLGSPKWTGQVPIPRGRVIMRN
          MPKVNGLIGDMMALRTCTVOANDPSSSEADMAKFPMSIALDPIATLAFNDSPTKGP
          NSELSRANWVDTDPNRTGLDFVEEDGFDERARVALDPIPMFVRRGKYIDVAG
          RSFRDPIBEKLPDELPELATKDFADHTTAPREVALRTLYEMRCADGAPMSKALP
          ALMTGVFYDDALAAAMDLCRWTLLEDEBGLARDPKLGIAQVAGRSADQVAKRVA
          VYAGLKNRAQLNGGFLDETITLGLERQLADSGITPAEKLSLIVMGANGDIERVFTD
          CAY"
BASE COUNT 2025 a 3680 c 3749 g 1734 t
ORIGIN

```

Query Match 44.8%; Score 750.2; DB 1; Length 11188;  
 Best Local Similarity 67.5%; Pred. No. 2.2e-147;  
 Matches 1122; Conservative 0; Mismatches 518; Indels 21; Gaps 4;

```

Db 1620 GGCATCGCGGTGGCGGCTTTCCGGCCATCCGCTGGGAGCGAGGCGCGGCGATT 1561
Qy 184 cagtaacatgcttgaaagcgcaagcgtctgacatgctgacatgctgacatgacg 243
Db 1560 CAGGTGTTCTCCGACACCGG-----CATGCGCGCCGATGGCATCTTCAAG 1513
Qy 244 aacgagacacccgttgaaagcgtatgatacctacatgatacttctgacatgcttctg 303
Db 1512 GGCAGACGCGGCTATACGCGCGGTGTCACGCGCCAGCGCCGATCATGATCTTTCATG 1453
Qy 304 ggtatcccgacatgcttgctgctgctgctgctgctgctgctgctgctgctgctgct 363
Db 1452 GTCATGCTCGGATGATGCGCGGCTGCGGCACTGCTTCTGATCATGATGCGGCG 1393
Qy 364 ccgataatgcttcctcccgatgataaacaactgcttcgctgctgctgctgctgctgct 423
Db 1392 CCGGACATGCGCTTCCGCGGATGACACATCTCTGCTTGGCTGCTGATCGCGCTGG 1333
Qy 424 gcgatgagcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 483
Db 1332 ATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1276
Qy 484 ggttggtctgctgacccgctgctgacccgctgacccgctgacccgctgacccgctg 540
Db 1275 GGTGGACGATCTATCCGCGCTGCTGACATCCGCTGACAAAGGCTCCGCGCATGACCTG 1216
Qy 541 ggaatttcgagctgacatgctgctgctgctgctgctgctgctgctgctgctgctgct 600
Db 1215 GCGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1156
Qy 601 aagacatcttgaacatgacgctgctgctgctgctgctgctgctgctgctgctgctg 660
Db 1155 ACCAAGATCTCATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1096
Qy 661 tggctgacttatacagcgtctgctgctgctgctgctgctgctgctgctgctgctgct 720
Db 1095 TGGTGGTGTGCTGATCACCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036
Qy 721 atcacacatgctgctgacgacgctgacgacgctgacgacgctgacgacgctgacgac 780
Db 1035 ATCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 976
Qy 781 ggtgacccgactctgacacacacacacacacacacacacacacacacacacacacac 840
Db 975 GCGCGACCGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 916
Qy 841 atcattcgcgcgcgttgacacacacacacacacacacacacacacacacacacacac 900
Db 915 CTGATCCTGCGGCGCTTGGCATCATGACACACATCTGCTGACCTTCTCCACAAACCG 856
Qy 901 gcttcgcttcaacgctgacgctgacgctgacgctgacgctgacgctgacgctgacgct 960
Db 855 GCTTCTGCGCTACCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 796
Qy 961 gtcgagcgacacacacacacacacacacacacacacacacacacacacacacacac 1020
Db 795 GTGTGGCGCACACATGATGACATGACATGACATGACATGACATGACATGACATGAC 736
Qy 1021 ctgagcaccatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1080
Db 735 GCGCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 676
Qy 675 ATGTGGGCGGCTTCCGCTCAGCTCAAGACGCCATGCTGTGGGCGATGCGGCTCATCTTC 616
Qy 1141 ctgttaccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1200
Db 615 CTTTCAAGCTGTGCGGCGGCTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556
Qy 1201 tatagacacacacacacacacacacacacacacacacacacacacacacacacac 1260

```

```

Db 555 CTGCACGACACCTACTACGTGGTGGCGCACTTCCACTACGTGCTGTGCGTGGCGCGGTG 496
Oy 1261 ttggcatttcgacgagatcttcaactttcaacgcgaagttctgggcccgcgttcccg 1320
Db 495 TTGCGCATCTTGGCGGGCTTCTACTACTGTTGAGCAAGATGTGGGCGTGAAGTACAAAC 436
Oy 1321 gaatggcgtgcaaaagctgacattcgacattcattcatcgctgacgaagtcacgtcttc 1380
Db 435 GAGTTCCTCGGCGGGCCACACTTGTGATCATGTTTCATGCGTGTGACACTGATCTTCTTC 376
Oy 1381 ccgcagcacttcctgagacgttaaggatagccgcgcgcgttacatcgactaccogaagcc 1440
Db 375 CCGCGACACTTCTCTGGCGCTGCAAGGCAAGCCGCGCTATGTCGACTACCTCGGACGCC 316
Oy 1441 ttgcgcgttgcaacaagctcctccatagtgctgccttcctgcctgcctccctgcgttc 1500
Db 315 TTACACCTGTGAACTACGTCTGACGGTCCGCTACATGATACCGTCTGTTGTCTGCGC 256
Oy 1501 ttcttcacgtgatacttctgtctatacgtctgttctgacgcgcgaagccgtccgaac 1560
Db 255 GTGTCATGCTGTGCTGATGTCAGGGCGCGGAT---CCGTCGCGCGCAAGCGCGACGAC 199
Oy 1561 ccgtggcgcgaattccgcgcgcgtcgtgaaatggaagcgtgcacaccgcctccgcgcac 1620
Db 198 CCGTGGGCGGAAGGCGCGCACACGCTGGAATGACTCTCTCTCGCGCGCGCGCTTCCAC 139
Oy 1621 acgttcgaacgcgtcccaagcgcctcgactgagacaagca 1661
Db 138 CAGTTCAGCAGACCGCGCGTGATCAAGGGCGACGACCA 98

RESULT 7
AE008010 10835 bp DNA BCT 14-AUG-2001
LOCUS Agrobacterium tumefaciens strain C58 circular chromosome, section 68 of 254 of the complete sequence.
ACCESSION AE008010 AE007869
VERSION AE008010.1 GI:15155740
KEYWORDS
SOURCE
ORGANISM Agrobacterium tumefaciens.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.
1 (bases 1 to 10835)
REFERENCE 1 Hinkle,G., Slater,S.C. and Goodner,B.
AUTHORS Complete Genome Sequence of Agrobacterium tumefaciens C58
TITLE (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10835)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
source
1.10835
Location/Qualifiers
/organism="Agrobacterium tumefaciens"
/strain="C58"
/db_xref="taxon:358"
/complement(118..642)
/gene="AGR_C_1396"
/complement(118..642)
/gene="AGR_C_1396"
/note="(AJ242592) hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1396p"
/protein_id="AAK86575.1"
/db_xref="GI:15155741"
/translation="MRLSPILRTFLAAGLAIIPAAALAAOQOPGPKSTHGAWSY
ICDKPAGASEGALMONVITADDPKGLSVITIKTADPKRSRIIRIAPLVLLKQDM
ELIYDNNNIGATYTRCFSEGCYIEVDIDELLVLRAGNAVPALRESVDQDRVIP
IELSGFAEGIDALP"

gene
988..1938
/gene="AGR_C_1397"
988..1938
/gene="AGR_C_1397"
/note="(AJ242592) cytochrome c oxidase subunit II"
/codon_start=1
/transl_table=11
/product="AGR_C_1397p"
/protein_id="AAK86576.1"
/db_xref="GI:15155742"
/translation="MDSASVIDDQTRREGCTVTRNRIYAALAGMTCLITAVGAHADOP
VHMOMGDEATPIHMEIRMEFOYTLMEFVPTLFLVALLITVAVKFHAAPKAPVSKT
SHNTAEVYVATPVLILFLAPSENLNADQEPNDLTKATATQMLMSYKTA
AEGAPLSPDSYLLDODPARAGKEDKARYPELLAVDNDMVYPCKTARLLTAAPTD
VTHAFAPFGKIDAVPGRRLNETFKPEKECLYTGCCSELCGKHAFPIAIRVSE
DOYNTVHAASADLNGANRALMASVDGAPRTVDVAAETN"
2549..3664
/gene="AGR_C_1399"
2549..3664
/gene="AGR_C_1399"
/note="CYTOCHROME C OXIDASE POLYPEPTIDE I (CYTOCHROME A3
SUBUNIT 1)"
/codon_start=1
/transl_table=11
/product="AGR_C_1399p"
/protein_id="AAK86577.1"
/db_xref="GI:15155743"
/translation="MDLAFSLHVGASSTIGAINFITTILNMRAPGWTLHKMPLFAW
SVLVAFLLLSLPLVLAGITMLLDNRNGTAFESPEGGDPIILYQHLEMFEGHEVY
ILILEPGILISHIVSTPSFKRPFVGLGNAVAVAIAGVFIWAHMHVTVGLSLAOR
YFEFATNVIAPVTGKIKIFSMITMGSGLTFSTPMWAIIGFLEFTVGSVTVOLANA
GLDRSLHDPTVYVAHGHVLSIGAVFAIFAGVYVMPKPTGYNTGNTNEFTGLKHPMFI
GVNLIFPQPHGLAGMPRRYIDYDPAIAGNMVSYSYIAVAVGIFLFEVWAFPA
KKRIAGNMPGSGATTLTMOUSSPPYHOMEDLPRI"
4054..5007
/gene="AGR_C_1402"
4054..5007
/gene="AGR_C_1402"
/note="(X89566) heme O synthase"
/codon_start=1
/transl_table=11
/product="AGR_C_1402p"
/protein_id="AAK86578.1"
/db_xref="GI:15155744"
/translation="MTVIDDDRMGAESSELSAGARDYFELKRPWLSLVFTAFAG
LVLAEGELNPLIGLAIICIAVGAGSALNWDYADIDAVMTTRAKRPIDSGRIAPR
EALAEGLTSAFVSIVILGAVNMFSAAGLAFIPFVAVYTMWLRSTPONTVIGGA
GAFPEMLGMACVTVGVSLSVILFLIIFLMTFAHMAVALFEMRDYGAVGIPMPNVA
GERSTKNOMIYAVVITAAVAAPFETGLASAGYGFPAVLVAIFVYCSQEDVRMPEGD
EKMLPAKKMPAVSVILFALFISGLADHFAPAKAVISGLV"
5167..5784
/gene="AGR_C_1406"
5167..5784
/gene="AGR_C_1406"
/note="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
/codon_start=1
/transl_table=11
/product="AGR_C_1406p"
/protein_id="AAK86579.1"
/db_xref="GI:15155745"
/translation="MANMEAEETARAGSORVNSRIIVLCVLFECMIGIMAAVAAPLYS
LECRVYNGTTOREYOYSDVILDKTIVTIDGANSGLNDPRVDKRVKPKIGETI
KVFETKATRSVATGTGAVFNVTPMEACATKRYKVCREFETTLQDQGTLEMPVFFV
DPDITARETYSVHTLITSYTFYPAKTEKPVASLPVKTESGSKL"
5846..6721
/gene="AGR_C_1407"
5846..6721
/gene="AGR_C_1407"
/note="(AJ242592) cytochrome c oxidase subunit III"
/codon_start=1
/transl_table=11
/product="AGR_C_1407p"
/protein_id="AAK86580.1"

```



```

/db_xref="GI:15155746"
/translation="MADTHOKNHDYHIIDPSPWPLLASIGAFIMTGGVCVMRTLSG
SPKLEGELANPMLFYIGLVIVYVYAMWADTIKEANESGSHTRVSLHLYGIMELI
ASEVWFPAWEHAYFADASLFPHEAIOASRLLEYTGQMPKCIETVIDPMHLDLYNTVIL
LISGCVTWAHHAHLHNDKGLISGLATLVALGVLFSTVOYEYTHAFEDKRNSTYGA
TFPMATGHEHGFVETVLLVCLFRAIAGFTPKQHPFBAAWYMHFVVDVWMLFE
FAIYIMGCGAPLHG"
/gene="AGR_C_1409"
6941..7369
/gene="AGR_C_1409"
6941..7369
/gene="AGR_C_1409"
6941..7369
/translation="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1409p"
/protein_id="AAK86581.1"
/db_xref="GI:15155747"
/translation="MYSRFGCQREHREVLMSSEPSNGNFAPVDPIKVLGKCCPKCGNGR
LEDFLTPKPCASGGLDYGFADAGDPAVEFMILVGLVYGMALWMDORPAPYMAH
VMLMPLFTVYSLVLRKLGIMIALQYRNNASEGRDRE"
7338..8114
/gene="AGR_C_1411"
7338..8114
/gene="AGR_C_1411"
7338..8114
/translation="surfeit 1"
/codon_start=1
/transl_table=11
/product="AGR_C_1411p"
/protein_id="AAK86582.1"
/db_xref="GI:15155748"
/translation="MQANAGLIVNDVSHMKAQRWWEAPLVLLALAILLGCTWQY
KRLYKEMALMADIEERRNAPATLSIDIAIVSGGEIYRKVRLSGTDHREIRFEA
THOQGGYIYPLFLADRIILFNRGVPEFEMKAARPDGOVGEVYITGLAAPL
VAKPSLLPDNDIAKNIFWKDLAAMASAEVPERIKLFDADANAPNPGMPQGGV
TIIIDPNHLOQYAITWYGLAALVIVAGFAFRNKAQGE"
8173..9219
/gene="AGR_C_1414"
8173..9219
/gene="AGR_C_1414"
8173..9219
/translation="LybB"
/codon_start=1
/transl_table=11
/product="AGR_C_1414p"
/protein_id="AAK86583.1"
/db_xref="GI:15155749"
/translation="MNIAVSKPPLTIRLCGPRGACADYRAIQIIVLAKAAGFVY
RHEIYHNRVYVELEKGAIVEELHEIPAEHREBPVESAQGVSEPDQOANLF
VIDATCPYASKYHKQAMRHQRHVILIGHAGHREIVGTGMOLEPGTVSLVETVEDA
GYEYEDRENLSFVYOTTLISVDDTNGAVIARLDERPPIAIOARADSTICATTPROAVK
QAPGCDLFIIVGAPNSNSKRLVEALNAGAKHSVLVORASEIDMNEIGDIRTVGLS
AGASAEVIVDEIIAFAKARFDITLDAVTEETHFLVRELRSIELTTDMFAVNG
NASNALPKATAGI"
9312..10280
/gene="AGR_C_1416"
9312..10280
/gene="AGR_C_1416"
9312..10280
/translation="HOMOSERINE KINASE (HK)"
/codon_start=1
/transl_table=11
/product="AGR_C_1416p"
/protein_id="AAK86584.1"
/db_xref="GI:15155750"
/translation="MAVYTDITDELDNELNLTQYDVGSLTSYKGAEGVENSFLLHTT
KDPILITLVEKVERNDLPFLGMQHLAAKGLSCPLPRKDGELBELSGRPAALI
SFLGEMWLRKPEAKHREYKALAMHLASEGFEIKRPALSDGMKYLMDSEERAD
EYKGLREIIRPEIDYLAHWMPKDLPAVITHADLPDQNVFELGDELSGIDYFACND
LLAYDSICLNAMCEPKDGAAYVTKKALLBEGYOSVRPSELEALPLLSGSLRFR
ETRLVDWLTTPAGALVYKADPLETLRKLRHRTIANVAEYGLAGE"
10277..10717
/gene="AGR_C_1417"
10277..10717
/gene="AGR_C_1417"
10277..10717
/translation="Ribonuclease H (E.C.3.1.26.4) Mutant with His 62"

```

```

Replaced By Pro (H62p)
/codon_start=1
/transl_table=11
/product="AGR_C_1417p"
/protein_id="AAK86585.1"
/db_xref="GI:15155751"
/translation="MKHVIDFTDGACSGNPGSGVAVLYRGTENEKLSGGEDATTNN
RMELAIISALNALKSPCEVDLYTDSAYVDGTYTKWIFMKKKGMKTADNKPVKNNEL
MOALEAEOERHRYTLHWKGHAGHPENRDELARKGMEPFRR"
BASE COUNT      2021 a 3194 c 3242 g 2378 t
ORIGIN

```

```

Query Match      44.2%; Score 740.4; DB 1; Length 10835;
Best Local Similarity 68.0%; Pred. No. 2.6e-145;
Matches 1109; Conservative 0; Mismatches 511; Indels 12; Gaps 5;

```

```

QY 13 gccattcaagccatgaccaccatgagaagaagcctcttcacgcgtggttcacg 72
DB 2025 GACGATCATTTGCGATGATGCCATGAGCCGGGCTTTTCGCCCTGCTCTGTC 2084
QY 73 accaaccaagaagatgctgctgctgctgctgctgctgctgctgctgctgctg 132
DB 2085 ACCAACCAAGAAGATGAGGACGCTGATCTTCCGATCATGCGGATATCATC 2144
QY 133 tccgctccttcacccgtctacatgctgctgctgctgctgctgctgctgctg 192
DB 2145 GCGCGCGGCGCTTTGGGTGCTGATGAGGTGAGCGGAGGAGCGGCGATCATCTTC 2204
QY 133 tgccttgaagcgacgctgctgctgctgctgctgctgctgctgctgctgctg 248
DB 2205 CACGCTTCGCTTCATGCTGATGCTGAGGCTTCGAGGCGCATGCCGCGATGAG 2264
QY 249 acacctgtggaagcgtacgtgctgctgctgctgctgctgctgctgctgctgctg 308
DB 2265 -CACATGTTCAACGCTGTTACGACGACGCGCGCTGATCATGATCTTTCATGCTAT 2323
QY 309 cccgcgcatgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 368
DB 2324 CCGCGCGATGATGCGCGGCTTTGGCAACTGATGATCCGATCATGATGCGCGCTGA 2383
QY 369 tatgctcctcccgctgctgctgctgctgctgctgctgctgctgctgctgctg 428
DB 2384 CATGGCTTTCCCGGCGCT- AACAAATATTTCTTCTGCTGATGCTTCCGCTTCATCTT 2442
QY 429 gggcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 488
DB 2443 GCTGCTCTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2502
QY 489 ggtctgtaacccgctgctgctgctgctgctgctgctgctgctgctgctgctg 545
DB 2503 GACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2562
QY 546 ttgcggttcaactgctgctgctgctgctgctgctgctgctgctgctgctgctg 605
DB 2563 CTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2622
QY 606 cttcttgaacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 665
DB 2623 CATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2682
QY 666 gatcttatacagctgctgctgctgctgctgctgctgctgctgctgctgctgctg 725
DB 2683 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2742
QY 726 catgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 785
DB 2743 CATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2802
QY 786 ccgcatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 845
DB 2803 TCCGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2862

```



QY	846	tcgcgcgcgcttcttgcaatcatcaacacaaatgcgtgctgctgacaccttctgcgaaaaagccgctct	905
Db	2863	CTTGCCCGGTTTCGGATCATATCAACCAACATCTGTCCACCTTCTCCAAAGACGGCTCTT	2922
QY	906	cgcgtaccctgcgcgaatggctctctatctgaatgctggtgcgaatcgcgtgctcttgagcttgcgtc	965
Db	2923	CGGTATCTCGGCGATAGGGCTTAACGCCATAGTCGCCCATGCGGCTGGGCTTATCTCTCTG	2987
QY	966	ggcgcacacaatgtaaacgcgttctglatgctgcgtgaacccaatccctactcaatgctgcgc	1021
Db	2983	GGCCACACACATGTATACACGGCGGCGCTGTGCTGCATGAAGACACACGCTACTGTGTTTCCG	3041
QY	1026	caccctgtgtatgcgcgttgcgcgaacggccatlaagaatctctcgtgatacgcgaagctgc	1087
Db	3043	CACCATGCTCATTCGGCGGTGCCGAGGGGATCAAGATCTTCTCTGATGATGCAGCATGTCG	3102
QY	1086	gggcgcgtcgcgtcttgaagtaacaaatgcgcgaatgcctcgtgcgccttgcgttatgtccgtgt	1145
Db	3103	GGGGGTTTCGGCTGACCTTTCTTCGACCCCGATAGGCTGGGCAATGGGCTTATCTTCTGTT	3165
QY	1146	caccgtgggtgtgtgtgacccggtatcgtgctgcgcgaagcggctctgacacgtgacatca	1205
Db	3163	CACGCTGGTGCTGCACCGGTGCAGCTGCCAATAGCCGGTCTGCAGCGTGTTCGCTGCA	3222
QY	1206	cgaacccatctaacgtgtgtgcgcgaactccaatcttgatgctgcgtgcgtgcgtctctgc	1265
Db	3223	CGACACCTATTACCTCGTGGCTACCTTCCACTACGTGCTGTCTGCTCGGGCCGCTTTGC	3283
QY	1266	gactctgcgcgctatcctaactcttatacabgcgcgaagtctctcgtgcgcgcgtcttccgcgaatg	1325
Db	3283	GATCTTTCGGCGGCTGCTATTAAGTCTGTTCCGGAAGATACACGGGCTACATGTACACGAGTT	3342
QY	1326	gagctgcgaacgtgcactctgcgaactcttctcaatcgcgtgcgaagctcaagctcttccgcga	1387
Db	3343	CATGGCGAAGCTGATTTCTTGGGTCATAGTTCACTGGCGTAACCTGTATCTTCTCCGCA	3402
QY	1386	gcaactctctgggaagctaaaggtatgcgcgcgcgttcaatcgaactaccgaacgtctgc	1445
Db	3403	GCACCTTCCTTGGTCTGCCGGGCAATGGCGGCGGCGCTCATCATGATTAACCGGATCTTACGC	3462
QY	1446	gctctggacaacaagtctgcctcctaattgctgcttcgtccctgccttcgcctctgctctctc	1505
Db	3463	CGGCTGGACATGGTATACGTCCTACGCTTCCATAT--CTGAGCCGTTGGCGGTGGCAT	3519
QY	1506	catcgtgatacttgcctctatacgtctgcttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1565
Db	3520	CTTCTCTTTCGGCGCTCTGGGAACCTTTCGCCAAGAAGCCCATGCGCGGTAAACAATCCCTG	3579
QY	1566	gggcgcgaatctgcgcgaatacgtgtaagtaagcgtctgcacatcacgcctcgcgcgaacgct	1625
Db	3580	GGGTGAAGTGTGCAGCAGCAGCTTGAATGGAGCGTGTTCGGCGCGGCTTACACACAGTGTG	3639
QY	1626	cgaagacgtgcgc 1637	
Db	3640	GGAGACAGCTTCC 3651	

RESULT	8			
LOCUS	AP003011			
DEFINITION	AP003011	346510 bp	DNA	BCT 15-MAY-2001
ACCESSION	AP003011	Mesorhizobium loti DNA, complete genome, section 18/21.		
VERSION	AP003011.2	GI:14026664		
KEYWORDS				
SOURCE				
		Mesorhizobium loti (strain:MAFF303099) DNA.		

## REFERENCE

Y. Kaneo, Y. Nakamura, Y. Sato, S. Asanuma, E. Kato, T. Sasamoto, S. Watanabe, A. Idesawa, K. Ishikawa, A. Kawashina, K. Kimura, T. Kishida, Y. Kiyokawa, C. Kohara, M. Matsumoto, M. Matsuo, A. Mochizuki, Y. Nakayama, S. Nakazaki, N. Shlimpo, S. Sugimoto, M.

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL
-------	---------	---------	-----------	---------	-------	---------

Takeuchi, C., Yamada, M. and Tani, S.  
Complete genome structure of the nitrogen-fixing symbiotic  
bacterium *Mesorhizobium loti*  
DNA Res. 7 (6), 331-338 (2000)  
21082930  
2 (bases 1 to 346510)  
Kaneko, T.  
Direct Submission  
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research, Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp)  
URL: <http://www.kazusa.or.jp/nihzabase/>  
Tel: 81-438-52-3955 (ex 2338), Fax: 81-438-52-3934  
On May 11, 2001 this sequence version replaced gi:11994986.

gene  
CDS

```

L..346510
/organism="Mesorhizobium loti"
/strain="MAFF303099"
/db_xref="taxon:381"
complement(242..1264)
/gene="ml7087"
complement(242..1264)
/gene="ml7087"
/note="similar to lipopolysaccharide core biosynthesis
glycosyl transferase lpsB"

```

gene  
CDS

/protein\_id="BAB53261.1"  
/db\_xref="GI:14026665"  
/translation="MAGGLACAVRLVSDVHLPRLLAIADRPGAIDRKTQNIERR  
FEIYQKRVHDTESDVAAADIVLFWFQIAKLPLEPSEIARSDDLVTIGTCS  
EGEEFPEGLATLRLRAARFVNNRKLHEHYAPLRTIVHTYTPNCDVDTFFCRRP  
VSGELIRVMASGILGNGHGHANRSGFHDVIEPVAAPAGRLTALIREQHRHNDDE  
RLDIYVAVSRECGPNCLPAACGFIPLTVTRGNRDELVDQDNCILPFDGNA  
DKLRLATPRLSRSMARMTETIRGMWRVQAANVAMHTSLITAGAAEPYR  
DVPQGT"  
complement(1228..2265)  
/gene="ml17088"  
complement(1228..2265)  
/gene="ml17088"  
/note="similar to lipopolysaccharide core biosynthesis  
glycosyl transferase LpsC"  
/codon\_start=1

gene  
CDS

gene  
CDS

IAIAPDRGGLVLSSEKAGEMSKPRKLAKSELGAAPENRPLPAK"  
3870..4556  
/gene="mlr7090"  
3870..4556

```
/gene="mlr7090"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53264.1"
/db_xref="GI:1402668"
/translation="MAGKAKTDSPLVAARMSDIAYERILEGLFEDRRVPAGATISON
ELSAIVDPAVAPLRDALVLEAEGLITTHPRSGIOFVPRGELTRATYOFISYERAA
VRVPAEGODEAVMANLEIRSHRLRVERDGLKQLEEMALFEGELHGAIVIALRNP
LIDSTYRRMHNTRLLRDLKRTTPIMVRLKEHLDILEACGTNRADAAEKLOAHQO
AAMORSIDLV"
/gene=".5922
/4684..5922
/gene="mlr7091"
/4684..5922
/codon_start=1
/transl_table=11
/product="sugar ABC transporter, periplasmic binding
protein"
/protein_id="BAB53265.1"
/db_xref="GI:1402669"
/translation="MQLKROFLMSALMAAGHTRCARASATSYVHHPTOSEMAS
LKITGLFOANADVVYQENIPNSIYAKYSSAVLAGRPDTAVIYERADLTAAE
GLIDJDRVNGKRGKAPENRMWGLSSNGAIYAVPAFVDMYTRADYEEEGLAG
PRNDEFELACRLTDPKSKRYAFAGMAGAGFYVIDVMEATGSPYKDGOPALDR
AAAIATFYSDDLPLEKRYVPSAPNDYRQIMEGFTQOTAMVHHGTSINLSAAM
KOGQFSTAPMAGPKAHVARYAVANGIMKEDNTEAMRWVSEMGQDALTLEGT
GYEPAASALODPRITGNPIYAAETLEFCRLPNPFCAAGWSENVNPAFQAVLIG
OSTPEQAVDRMIOGLEAALR"
/gene=".6932
/5979..6932
/gene="mlr7093"
/5979..6932
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="BAB53266.1"
/db_xref="GI:1402670"
/translation="MLSHRGSMTSSSGYRPLRALGESETRYMYMLLLESLVAILL
VIAPYLVGMATSFREMRITRPGIGWGFYKHYAAMSDPYEMLAKTNYVAAVY
LELAGLLAATNLNDIDPCTKTEFVILITPLPVVAGHMAILLDRLGVINDIIV
RAGVLOTYKAMPADNPNTALATVLEAHMGPFFRLLLAGLKSIPEDLYKAAAVDGA
GVSAQRLITPLMLTIVIAAAILVIGLVNSPDLILLTGSGEDATQVLSYAFOT
AKKENFGASALSVMEVILMIFATAYIRLSRTKGGDD"
/gene=".7770
/6925..7770
/gene="mlr7094"
/6925..7770
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="BAB53267.1"
/db_xref="GI:1402671"
/translation="MIESRRRRVIANIATYVLTLSMFCGPMHMFJLSKLEAD
IYVQIMQYIPRRVEDNYVAIVTOSGPEPLVANSVVFVLTIVATGTLAAVAFSR
FAFRGTQMLRGLYLVRMFPVAMMIIPLVVWNGVGLVDSRGLAATSPLLPLEVW
LMKGFDSAPKELESARIDGSTRGAMVRIYMPIARGLVASVFIATAAMNEFLFA
LMITGGGSRTPVGIQLMWGEFQIPWGLANGILSTILPVIVLEPAIVQRTWVQCLTA
GAIKG"
/gene=".8860
/7781..8860
/gene="mlr7095"
/7781..8860
/gene="mlr7095"
/codon_start=1
/transl_table=11
/product="sugar ABC transporter, ATP-binding protein"
/protein_id="BAB53268.1"
/db_xref="GI:1402672"
/translation="MATLSIKDVRRKSYGSAVQLHGVDELDEGFLVLLGPGCGCKST
LNIITAGDTSQDGLIGRSVVDLAPDRNIAVPOSAVLYPTMSYERNIGFCELM
RGVADQRRKRAVDAAARLLOISHLDRRANISGGOBVAAGRAIVRPELLEDP
LSMLDKLVEKTEIKRLHERIGTSIYVTTDOIEAMTLAKVAVMMGCHTQOPADP
RTYERPAVFAVSTIGSPAMNPLPERLVSATGVFAASGVSAIÖAKGRTVDSERP
```

```
VLIGRPEELKAVGRDEASVGVIDVEPTGPEWTVYQVQDSIIARLPFRAGRN
EPFLAVDPASVNLFDPGSEHRIIDT"
/gene=".9082
/8654..9082
/gene="msr7097"
/8654..9082
/gene="msr7097"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53269.1"
/db_xref="GI:1402673"
/translation="MDAKALLAORGFGRHPRMSRSHHRCRRDRRRLQRLMAVAEAY
AGCYRLYGEINQMIIPMWGRIAQART"
/gene=".10692
/9079..10692
/gene="mlr7098"
/9079..10692
/gene="mlr7098"
/codon_start=1
/transl_table=11
/product="GMC-type oxidoreductase"
/protein_id="BAB53270.1"
/db_xref="GI:1402674"
/translation="MSPTSPFGDYDYVIVGASAGCVLIANRLSADPRNRLVLLFAGGS
DRHWVDIPIGITFCMKNPTPTDMMKTEPBAAGINRSLNTPRCKVLGCGSSVNGMYM
RGAADYDGRQAGNGVMSMDVLYPFRKSEDIHGKSMHSGGEWREORLSMPT
LDFAERDAEELGIPRTGDFNTGNEGSEYFVNOQLGVMWNTSKAFRLRQIARQNDLV
VTGAETRELEFEDGRVYGLLEFGDGLHRVSAQOTILAGAINSPILBELSGVRDL
LSSIGYVRRALPGVGENLDQIOTRIVEVNAATLNEBYHNLISRAMGEYELAKR
SGPLSMASQDGLFARSADRLATPDLETHVQVSTORLEPLLRPAVAVSCNLEPE
SRGVYHMASSDPREARTIRPNYITACGDSVAASLRHVRSLMKARAVALRPAPEMLP
GTOTDSEDELIRKAGDITGITTFHPVGTCKMGSMAVVDRLRVHGLAQLRVYDASIM
PTIVSGNTNSPVVMIERKAEMITLDPRR"
/gene=".11801
/10929..11801
/gene="mlr7099"
/10929..11801
/gene="mlr7099"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="BAB53271.1"
Query Match 43.0%; Score 719.6; DB 1; Length 346510;
Best Local Similarity 65.7%; Pred. No. 3.2e-141;
Matches 1097; Conservative 0; Mismatches 364; Indels 9; Gaps 3;
QY 1 atggcagaagccgcgcattcacgcgcatgaccacatgagaagcaagcctcttcacgcgc 60
|||||
DB 337162 ATGCCAGACGCTGCGACGCTACGACAGCGCCACGACCAAGCCTTATCATGCGTGGTGGC 337221
QY 61 tggctacatgctgaccacacacaaagacatcggtctgctataccttgaagcgtggtgt 120
|||||
DB 337222 TGGGCTCTATTGACCAACACAGACATCGGACGCTCTACCTGATCTTCGCGATCATG 337281
QY 121 gttgttcatttcctgcttcgctacacgctcacaagcgccttgagcgtgatgacgggt 180
|||||
DB 337282 GCGGCGATCATCGGCGCGCGCTGTGCGTCCGACATCCGATGAGATTGCAAGACCTGGC 337341
QY 181 gtccagtaacatgctcgtgaagcgcaagctgacatcgagcgtgagctgcagacatgac 240
|||||
DB 337342 ATCCAGATCTTTCAGCGGCTGTGGCCAGATGCTACGCGCATGAACGGTGAGCGCGATT 337401
QY 241 gcgaacggaca---ccgttgaagcgtcatggttaacctacatgatatcgaatgatttc 297
|||||
DB 337402 GATGCGCGCAGACAGCATGTACATGCCCTTCGACCGCGCATGCGCTGATCATGCTTC 337461
QY 298 ttgttggtatcccgccgttgcgttggttgaatcctcgaatgcgcgtgcaatc 357
|||||
DB 337462 TTCATGCTATGCGCGGCTGATTCGCGGCTTCGCCAATGATGATGCTGATCATGATC 337521
QY 358 ggcgcctcgagatatgacctcccgatgaaacacatgctgcttgcgttgcattgacc 417
|||||
DB 337522 GCGCGCGCGCGACATGCGGTTCGCCGCGCATGAACACATCTCTGCTGCTGCGCGCG 337581
```



```

/translaction="MATASANDHADDHGHANHPGKREAFYSTNRKIDTMYLVFA
IVAGILIGMISGIRAEIMYDGOIJEVDEPAHNYNEATPHGLIMIEFPMMPAMIGFEN
MFPVILIGAMPAPRNNISITPMILPAFALLISSEMEGPGATGAGTGTATLAPIS
TSIHGPAVDSTISILHLAGASSITIGAINFTTTIFPMNAPRMTNKNPLFWMSLVIV
FLLLALPVLGALATMLLTDNRFGTAVYMAVAIGIGVMAHMYITGVMSAAYVAAT
FGMISHVISTESKPIFGYLGMAVMAIGIGVMAHMYITGVMSAAYVAAT
MVIAPVGVKFSMTATMVGSIIEKFPMLFAIGIFLFTVGVGVLAAGVDRVM
OETYYVIAHNFVMIAGVAFISGMWYPMKSGMYNENYIKLHFMVMTIGANILE
PFOHFLGCGMRMVDYPDAPFAGNIEISSYGFATFGRVILIPLYGLDAPARRQQA
DNPMEGATULEMTLISPPPHQSTLTKVQ"
314..4303
/gene="cox"
/gene="cox"
3314..4303
/gene="cox"
/codon_start=1
/transl_table=11
/product="heme O synthase"
/protein_id="CAA61745.1"
/db_xref="GI:1177575"
/db_xref="SPTREMBL:O51322"
/translaction="MCAAHSCAHPRTKAGSTVSVVDOKAVELPRISFAGVADYFALL
KRVMSLVITFALVGLMATAFGHVPVLCFIALICIAVAGASGALNMLBEDIVIMS
RTANRPIDPRGRTREGAMFGLTISFVSMTDGLVMTAGGLAFITFEVYVITMG
LKRTAONIVIGAGALPVPVMAAAGSLSEPELFLIIFETPPEFWALAFERS
DDYARAGVPMLEPVVAGPADATRLQILLYIVLVAIAAAMPWLGPEWIMYGVSLTIGAG
MLVAIEVYRHRHRSOALRATRLPAFISIFLYFLFVLLDVAKAVAPLIW"
BASE COUNT      749 a      1252 c      1415 g      964 t
ORIGIN
Query Match      42.4%; Score 709.2; DB 1; Length 4380;
Best Local Similarity 70.5%; Pred. No. 1e-138;
Matches 977; Conservative 0; Mismatches 403; Indels 6; Gaps 2;
QY 259 aaagtcattgctccaccatggtatctgagatgctcttgtagtcccccagatg 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1910 AACGTTTTCGACCAATCAGCGCTGATCATCTTCTCAAGTGATGCTCGGATG 1969
QY 319 ttccggtgtttgttaactatctgactgacatcgacgctccgagatagtcctc 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1970 ATCGGGGGCTTGGCAACTGCTGCGCTGATGATCGCGCGCCGCGCAATGCGCTTC 2029
QY 379 ccgagatagaacaactgtctcgtgctgctatgctcggtgacgagatggcgctg 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2030 CCGCGATGAAACAATCTGCTGCTGCTGCTGCGCGCGCTTCGCTGCTGCTGATC 2089
QY 439 tccgctcttcacacgagctgacgctgacgctgctgctgctgctgctgctgac 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2090 TCCTCTGTTCAATGAGGGGTAGCCGGCGCACCGGAGCGGCGCTGACGCTGTAT 2149
QY 499 ccgacgctgtcgaaccgctg---aagctgtatctgactgagaccttcgcggtt 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2150 GCGCGCTGTGACGAGCGGTGATCCGGGTCGCGGCTGACATTCATTCGTGATC 2209
QY 556 caactgtcgggtgctcctcctcgtatcagtgagtgagacatgatacagactcttgaac 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2210 CACCTTGGCGGGCGCTCTCTGATCTCGGGGATCACTCAATCAACAGATCTTCAAC 2269
QY 616 atgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2270 ATGCGCGCGCTTGGCATGACCATGACCAAGATCCGCTGTCTGCTGATGATCCGTGTC 2329
QY 676 acgagcttgatgactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2330 ACCGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2389
QY 736 accgacgcgaactctggaacgactcttcaatccgcgtcgtcgtcgtcgtcgtcgtcgt 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2390 ACCGACCGTAACTTCGACACGCTTCTTGCGCGCGCATGCGCGCGCATCCGCGTCTG 2449
QY 796 taccacaacacccctggtctcttggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2450 TTCCAGCATCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2509

```

```

QY 856 ttggacatcaacgacatgctgctgacacttcgcgaacaaagccgctcctgctgactg 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2510 TTCCGACATGATCAGCAGATGATCTCGACGCTTCGCAAGAACCGGATCTTCGGATATCTG 2569
QY 916 ccgattgtctatgcaatgtaggcaacgcggtgtcttgagcttctgctcgtcgtcgcacac 975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2570 GCGATGCGCTTCAGCGCATGCTGCGCATGCGCGCGCGCTTCGCTGCTGCTGCGCGCACAC 2629
QY 976 atgtaacgccttgtagtgcctcgtacccagcaatccactcaatgcgcgcacacagtagt 1035
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2630 ATGTACACGCTCGCATGCTGCTCTCGCGGACGCTTACTTCTGCGCGGACCATGTGTC 2689
QY 1036 atcgcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2690 ATCGGCGTCCGCGACGCGCGGTAGAGTGTCTGCTGATCGCATGCGCATGCTGCGCGCTCG 2749
QY 1096 gttgagttcaaatcgcgcgaatgctcgtgagccttggtcttattgctcgttccacgtagt 1155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2750 ATCGAGTTCAAGACGCGCATGCTGCTGCGCATGCGCGCTTCTGCTGCTGCTGCGCGCTCG 2809
QY 1156 ggtgtagccggtatcgtgctcgtcgcgaacgaggtctggaacggtgacatacagacactat 1215
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2810 GCGGTGACCGCGCTTGTCTGCGCAATGCGCGCGCTGATCGCGCATGCAAGAACCTTAC 2869
QY 1216 taagtgtgagcgacactccatattgtagtgcgtgagtgagtgagatccttgacatcgcgc 1275
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2870 TACGTCATCGCGCACTTCTCAGTGTGATGCGCATGCGCGCGGTTCGCGATCTCTCG 2929
QY 1276 ggtatctacttctatcagtcgcgaagtctcgtgacgcttcccggaatgaggtcgaag 1335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2930 GCGTGTACTCTACTGTTTCCGAAAGATGTCGGGTACATGTAACAAGAACCATTCGCAAG 2989
QY 1336 ctgcactctgacacttctcatcgtcgtcgcgaacgctcaagcttctccgcgcgcacttctg 1395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2990 CTGCACCTTCTGCGTATGTTTATCGCGCGCAACATTCGTTCTTCGCGACCATCTCTC 3049
QY 1396 ggaagtcaggtatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3050 GGTCTGCAAGGCGATGATGCGCGCGCATGCGGTGACATTCGCGATTCGCGCGGTGAGAC 3109
QY 1456 aaagctcgtctatgagtgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1515
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3110 GAGATCTCTCTTACGCGCGCTTCTATCAGGCGCTTCGCGGTGATCTCTCTACGCGC 3169
QY 1516 ttgtctatacgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1575
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3170 CTGCTGACGCG---CTTCGCGCGCAAGCAGCAGCGCGCTGACAAATCCGTGGGTGAAGGC 3226
QY 1576 gccgatacgcctggaatggaacgtgcatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1635
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3227 GCGAGACGCGTGAATGAGACGCTGCTCACCGCGCTCGTTCGTCACAGTCTCCACCCTG 3286
QY 1636 ccgaag 1641
    |||||
Db 3287 CCGAAG 3292

```

```

RESULT 10
BUCOXAG 1830 bp DNA BCT 12-SEP-1993
LOCUS B.japonicum coxa gene for cytochrome aa(3) subunit I (BC 1.9.3.1).
DEFINITION X54318
ACCESSION X54318.1 GI:39505
VERSION coxa gene, cytochrome aa(3) subunit I.
KEYWORDS Bradyrhizobium japonicum.
SOURCE Bradyrhizobium japonicum.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Bradyrhizobium group; Bradyrhizobium.
REFERENCE 1 (bases 1 to 1830)
AUTHORS Gabel,C.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1990) Gabel C., Johns Hopkins University, Dept.

```



Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium. 1 (bases 1 to 200)

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	FEATURES
2.	(bases 1 to 2000)				
	Bott, M., Bolliger, M. and Henneke, H.	Genetic analysis of the cytochrome c-a33 branch of the	Bradyrhizobium japonicum respiratory chain	Mol. Microbiol. 4 (12), 2147-2157 (1990)	
				91211625	Location/Qualifiers

mat peptide	1854	.	>2000	
BASE COUNT	304	a	669	c
ORIGIN			614	g
				413
				t

Query Match	Similarity	42.3%	Score 708.8	DB 1	Length 2000
Best Local	Similarity	70.3%	Pred. No. 1.4e-138		
Matches	979	Conservative	0	Mismatches	407
				Indels	6
				Gaps	2
QY	250	cacctctggaacgtcatgttctactacacatglatctatgatagttctcttggtatc	309		
DB	300	cacaccttaccacacgtcttctgtagacatccccacggcgctatcatgattccttcatgcatg	359		
QY	310	ccgcgatctgtctgtgtgttcttgtaactatctcgatgcgcgtgcaaatggcgtccgat	369		
DB	360	ccggcgagatgattcgccggcgttcggcaactctggttcgtcccgctggtatcgccggccgcac	419		
QY	370	atgagctctccgcgatatgacaacccgtgatctctgacglttcaatccggtacccgagat	429		
DB	420	atggcgcttcccgcgcttgacacacatctctgattctgctctgccggcctcttccggcgtt	479		

[illegible]

QY	1567	gagcaatccgcgaatacgcctggaatgacgcctgcacacgcctccgcgcacacacgcttc	1626
DB	1617	GGTGGCGGGGACACACACGCTGAGTGAGACGTCGCCCTCCCGCGCGCTTCATCGATTG	1676
QY	1627	gaacgcctgcgc	1638
DB	1677	GAAAGCTGCTGCC	1688
RESULT	12		
LOCUS	BUA242592		
DEFINITION	Bradyrhizobium japonicum coxB, coxA, coxE, coxF, coxG, coxC, shbI genes, partial t1d gene and ORFs 133 and 177.		
ACCESSION	AJ242592		
VERSION	AJ242592.1		
KEYWORDS	GI:6006409 assembly protein; coxA gene; coxB gene; coxC gene; coxE gene; coxF gene; coxG gene; cytochrome c oxidase subunit I; cytochrome c oxidase subunit II; cytochrome c oxidase subunit III; heme A synthase; Heme O synthase; hypothetical protein; ORF133; protoheme IX farnesyltransferase; shbI gene; SUR1 homolog; t1d gene; t1dd homolog.		
SOURCE	Bradyrhizobium japonicum.		
ORGANISM	Bradyrhizobium japonicum Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium.		
REFERENCE	1 (bases 1 to 8121)		
AUTHORS	Rossmann, R., Lofler, H., Rossi, P. and Hennecke, H.		
TITLE	Factors involved in biogenesis of active cytochrome aa3 encoded by the coxBaFVC gene cluster from Bradyrhizobium japonicum		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 8121)		
AUTHORS	Rossmann, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-MAY-1999) Rossmann R., Eidgenoessische Technische Hochschule Zuerich, Mikrobiologisches Institut, Schmelzbergstrasse 7, CH-8092 Zuerich, SWITZERLAND		
FEATURES	Location/Qualifiers		
source	1..8121		
gene	/organism="Bradyrhizobium japonicum"		
gene	/strain="110spc4"		
gene	/db_xref="taxon:375"		
gene	complement(1..440)		
gene	/gene="t1dd"		
gene	complement(<1..440)		
gene	/gene="t1dd"		
gene	/function="putative role in regulation of carbon storage"		
gene	/codon_start=1		
gene	/transl_table=1		
gene	/product="t1dd homolog"		
gene	/protein_id="CAB56816.1"		
gene	/db_xref="GI:6006410"		
gene	/translation="MTNPDTTSLIDRANLIDRDYVRHNEFARGLAGADGELFLEYSTQFSLAMDNGRIKQATYDTSGGFLRAVKDAGCAXSDDSYSPALIRADAAVAAYVGGYSYSGSFAPPHITNVRILKGDNDPLDARGFTFKYKLLAEIDYLRKDP"		
gene	complement(542..1075)		
gene	/note="ORF177"		
gene	/codon_start=1		
gene	/transl_table=1		
gene	/product="hypothetical protein"		
gene	/protein_id="CAB56817.1"		
gene	/db_xref="GI:6006411"		
gene	/translation="WGLSRKARDSSLLNLIGLLAALLALPLSLAHOCAVRSVHGDMWICRTPPGAAQAEQALIQSVAEDRSNAGLIVLTKADOKSRIMRVAAPGLVLLPSGGLNLIDQDVRAGFVCLPNGCVAAEYVMDKLLGGLRTATATFIIPETPEGYLGFPFLSLNGLEGGYDKLP"		
gene	1523..2362		
gene	/gene="coxB"		
gene	1523..1612		
gene	/gene="coxB"		

1523. .2362  
/gene="coxB"  
/EC\_number="1.9.3.1"  
/function="cytochrome c oxidation (respiration)"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental  
/product="cytochrome c oxidase subunit II"  
/protein\_id="CAB56818.1"  
/db\_xref="GI:6006412"  
/translation="MRMSGRYGRHLGLAAVGLTATGSAFAELGQAPAEWMTLDOO  
SSSPMDNIVMHNFLFVLAITLFEVALLVYVYKFNANAPVSPSTINTLILEVA  
WTLVLYLVGTSVSPFLLEIDVPRADLTIKATGQWYMSAYPNQKPEPDSLKL  
ADDDPRLLGNENMNVVYKNTYRQYGVADYIHAFLPAQYKIDALPGLNETWORK  
AKTGMFGQCCSELGKDHAEMPIAIRVEDEKESAWETAKKFPASGSGTGYASAAG  
PIQ"

gene 2451. .4076  
/gene="coxA"  
2451. .4076  
/gene="coxA"  
/EC\_number="1.9.3.1"  
/function="cytochrome c oxidation (respiration)"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental  
/product="cytochrome c oxidase subunit I"  
/protein\_id="CAB56819.1"  
/db\_xref="GI:6006413"  
/translation="MATSAAHGDAQDHDEHNPATGMRRVYSTNHKDIGTMYLL  
GAVIYGAASAIIRALEMYPGQIFHEHTYNAVPTSGLIIFPMWPAHIGFGE  
EMVPLMIGADMAFPRMNNISFWLLPASFGILMSTFEGEGANGVAGMTWYPA  
LSSSHPPADVAFILSLHLAGASSILAIINPTTIFMRAPGMLTKMPLFEVSIIVL  
TYELLSTPLVLAGITMLDLNRGTGTFEAPDGGGDPVLPQHLFMRGHEVYIILIT  
PGFGISQIVTSFESKRPYEGYLGAAYAAVAGIGTGYVAHNTVTGMSATQAPYA  
ATMTAATPCTKRTFSWATIMWGGSTIEEPAPIMAVGILFTVGQGVGVLANGVDR  
VLQETVYVAHNVHVLISGAVFAIPAGWYVFPKMTGYMNETLAKHFWPTFGVMD  
VEEPFHGLGSGMPRRYVDPDAFAGNMLVSSVSYISGFGVILFLCYVIDAEKAKPAP  
AGDNWGAAGATTLTPSPPPHGFVLPVQ"

gene 4137. .5081  
/gene="coxE"  
4137. .5081  
/gene="coxE"  
/function="proteome IX farnesyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="putative heme O synthase"  
/protein\_id="CAB56820.1"  
/db\_xref="GI:6006414"  
/translation="MSVLQDNAVDIRPRISEAVGDIYALIKPRVMSLVITFALVGA  
MARGHSHVLAITSLLTAVGASGALNMLEDDIDAKMRTANRPIPRKIRTPREA  
MTFGVLAFEFVMTIGILVMYIAGALLAFITFEVYVITYMLKRMTAQNIYVGAAGA  
LPYVAAVAAGVTDVEPDLLEFALIFEWTPRHFWALIFRSDDYARAGIPLPVPAP  
DRTKILLYTYIIVAAVAPMALGVPAVYGVVSLTIGGMVLAINVMYRRRROS  
LRATKRIAFSLITLVAFATLLAEVFRALAPRAGA"

gene 5078. .5251  
/gene="coxF"  
5078. .5251  
/gene="coxF"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAB56821.1"  
/db\_xref="GI:6006415"  
/translation="MIAVADKPEPDGIVLTLEOKSSRRORSIALTALGVLVFFAV  
TMVGPALVVRPM"

gene 5257. .5901  
/gene="coxG"  
5257. .5901  
/gene="coxG"  
/function="biogenesis of cytochrome c oxidase"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental

CDS



```

/product="putative heme A synthase or assembly protein"
/protein_id="CAB56822.1"
/db_xref="GI:6006416"
/translation="MDHEPTISRODSRTARKGVGKGLRDVIVASIGGVAVIWMGAS
YAAPVWMCRAFGTNGTVOVATPAATPGIARIKAVRDNAPGIPMKFEFEQSE
IEVNGOVTVVTVNQARTAGQARTIAPLVGYSFKINCFCEETDYMAGK
REMPVYVDSIADHENDLSITLSYTFYVKDPVVKPLASGEGDKRGNL"
5981..6877
/gene="cox"
5981..6877
/gene="cox"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/product="cytochrome c oxidase subunit III"
/protein_id="CAB56823.1"
/db_xref="GI:6006417"
/translation="MATAQGNHNDYHLVPSRMPAVGSLAFIMAVGAIAMMHMFSA
APIVGVGVLYTMASMGSDVKEADYKDDHETVQLHRRGYILEIASEVMFVA
MFAVFNALPFPADVAHATRDVAFEGGIGTAAGCAVGTWPGIETEDWHLPLN
TLILTSQTYTMHALLENDROGKGLITLVYGLPFCVQAYEYSHAAPFAGN
VGAIFEMATGPHGHVLTGTFLLVCLPRAIAGHTPTKQHLGEFPAAMYHNPVYVN
LEFLCIYVMGHAETMAHGAH"
6975..7376
/feature="ORF133"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56824.1"
/db_xref="GI:6006418"
/translation="MNDTAGPRKSEPTVQLQSNRSLACKSPRCGGKLYAGFTLAPS
CDRCIDYAFIDTGDPAIFILMAGIIVACALLIVEKYQRPVYLHVLMLPLIAT
TLPLIRAKSLLIALQFNHKAQRLVDRAK"
7199..8121
/gene="shb1"
7199..8121
/gene="shb1"
/function="cytochrome a3 biogenesis"
/codon_start=1
/transl_table=11
/product="SUR1 homolog"
/protein_id="CAB56825.1"
/db_xref="GI:6006419"
/translation="MRADRGQVPAVLAACRAVAADPRHIALATARDEVAADRRAI
PQSGAGPRGRMNETARKPRVAGFALFTLLAAEVALGWMOLQRTAKHIELIA
IHERIAAPRIALPPAOMVALPARDEFERFETFAFPAAMYSSGSAVRKQASG
PGTMAFAPRIALPSGEMVYVTDAGFVENTMDRSEVRAVKIYVGGPVALTGLYRREP
PGMTLPAESRDRKLMFVRDHVAIASALGKTYAPFYIDLEQAPANGIPREPDLVHL
KDDHLOYAVTWALAGAVLIAGVWVRGRS"
BASE COUNT 1282 a 2639 c 2619 g 1581 t
ORIGIN
Query Match 42.3% Score 708.8; DB 1; Length 8121;
Best Local Similarity 70.3%; Pred. No. 1.1e-138;
Matches 979; Conservative 0; Mismatches 407; Indels 6; Gaps 2;

```

```

QY 490 gttctgacccgcgcgtctgca-----ccgcgaagctgtgctatctgacatgacatcgcgatt 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2916 ACCATTGACGTCGCGGCTGTGAGCTCCGGCCATCCGGCCAGGAGTCGATCCGGATC 2975
QY 547 ttcgcggttcaactgttcggtgtcctctcgatcaatggcgcgatcaatgacgaagc 606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2976 CTGTGCGTGCATCTGCGGGGGGCTCTGCTGCTCTGCGGCCCAACATTCATCACACG 3035
QY 607 tcttgaacatgacgccccgcgcgcatgacgtgacaaagtaccgctgtctcgtgtcg 666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3036 ATCTTCAACATGCGCGCGCCGGGCATACCTGACAAAGATCCGCTGTTCGTGTGTCG 3095
QY 667 attcttaacgagcttggatctctgctgcgcgtcgcggttctgagctgtgtgacatcac 726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3096 ATCTGCGTGCAGCTGTCTCTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3155
QY 727 atgctgtgacgcgaacglaaactctgcaagcaacctcttcaatcctgtcgcgcggtgac 786
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3156 ATGCTGCTCACCGACCGCAATTTGCGACGACCTTCTTGCGGCCGACGCGCGCGCGAC 3215
QY 787 ccgattctgtacaaacatctctgtgttcttgggacccgcgaagtgtacatcatc 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3216 CCGTGCTGTTCACACACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3275
QY 847 ctgcccgcgttggcatcatcagcatgtcgtgacacctctcgaaaagccggtctc 906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3276 CTGCCGCGCTTCGGCATGATCAGCCAGATCGTCTCCACCTTCGCGCAAGCCCGTGTTC 3335
QY 907 gttaccctgcgcgaatgtatcgaatgtagcaatgtagtgtgttcggccttctgtcttg 966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3336 GGCATATCTGCGCATGCGCTACCGCATGCTGCGCATGCGCGCATGCGCTTCTGCTGTG 3395
QY 967 ggcacacacatgataccgcttgtatgtcgtcgtgacccgaacatctaacttaatgtcgtgc 1026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3396 GCGCACACATGTAACAGCGTGGCATGCTGCGGAGACGAGGCTATTCTGTCGCGCGG 3455
QY 1027 accatgtgtatcgcgcgtgcgcgaacggaatgaatctctctgtgtgacgcgaagatgtg 1086
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3456 ACGATGATGATCGCGCTTCGACCGCGCGTGAAGATCTTCTGCTGATCGCACATGTGG 3515
QY 1087 ggcgcgtgcgtgtgagtttcaaaatgcgcgaatgctctgagccttctgtatgtctctc 1146
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3516 GCGGCTGTGATCGAATTCGCGCGCGCATGATCTGCGCGCTGCGCTTCTCTGTCTC 3575
QY 1147 accgtggtgtgtgtgacacggtatcgtcgtgcgcgaagcggtgtctggaacgtgtacatcac 1206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3576 ACGTTCGCGGCGGTACCGGCGGTGCTGCTGCGGAACGCGCGGTGACCGCGGTCTCCAG 3635
QY 1207 gaacacctaacgtgtgtgcgcaccttcaatagtgtatgtcgcgtgtgtgtgtgtgtgtg 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3636 GAGACCTACTAGTGTGTCGCGCATCTCCACATGATGCTGTGCTGCGTGCAGTGTTCGCG 3695
QY 1267 attctgcgcggtatctacttcttaccatgacgaagtctcgcgcgcgcgtcttccggaatg 1326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3696 ATCTTTCGCGCGTGTGCTACTACTGTTCCGGAAGATGACGGGCTACATTCACAGACAGC 3755
QY 1327 gctgcaaacgtgcaccttctggaaccttcttcaatcgtgtcgaagtcgaatgtcttccgcgag 1386
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3756 CTGCGGAAGCGCACCTTCTGGGTACCTTTCATCGGCGTGAACGTGTGTGTTCTTCCCGAG 3815
QY 1387 cacttccctggaagcgtcgaagtatgcccgcgcgcgttaccatgcgaacttccgaagaccttcgcg 1446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3816 CACTTCTCTCGGCTCTGCGGCGATGCGCGCGCTACGTCAGTACCCGAGCGCTTGCC 3875
QY 1447 ctgtgaaacaagaatctgtctctatgtgttctcgtgcgtcgcgcgtctgtctcttctc 1506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3876 GCGTGAACCTGTGTCTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3935
QY 1507 atcgtgatacttcttatacgtgtgtgtgtgcgcgcgcgcgaagaccggtcgaacccggtgtg 1566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3936 CTTATTTGCGGTATGATCATCC---CTTTCGGAAGAAAGTGTGCGGCTGCGGACAAATCCGTGG 3992

```



QY	1567	ggcgaatccgcgcgaatcgtggaatgagacgctgcacatcacgcctccgagcccaacgltc	1626
DB	3993	ggttcccgccgacacacacgcttgagatgacgctgacctccgacgaccttcacatcattcc	4052
QY	1627	gaacacgtgcgc	1638
DB	4053	gaactgctgcgc	4064
RESULT	13		
LOCUS	BJU33883/C		
DEFINITION	BJU33883 31495 bp DNA BCT 10-AUG-2000		
	Bradyrhizobium japonicum putative epoxide hydrolase EphB (ephB), putative stress-induced protein Ohr (ohr), putative transmembrane transcrip		
	two-component system regulator protein Tirr (tirr), putative (tcsr), integral inner membrane metabolite transport protein MtbA (mtbA), orf157, and orf235 genes, complete cds; Fo ATPase operon, complete sequence; probable acetyltransferase Pact (pact), mitochondrial processing peptidase-like protein Mpp (mpp), threonine synthase Thrc (thrc), two-component system transcriptional regulator Elmr (elmr), putative two-component system sensor histidine kinase Elms (elms), SUR1-like protein (shb1), and orf133 genes, complete cds; cox operon, complete sequence; orf177, rldB (rldB), signal peptidase Sips (sips), and orf375 genes, complete cds; and putative glutathione synthetase GshA (gshA) gene, partial cds.		
ACCESSION	U33883		
KEYWORDS	U33883.2 GI:8708897		
SOURCE	Bradyrhizobium japonicum.		
ORGANISM	Bradyrhizobium japonicum Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium.		
REFERENCE	1 (bases 1 to 31495)		
AUTHORS	Muller,P., Ahrens,K., Keller,T. and Klauke,A.		
TITLE	A TrpHpa insertion within the Bradyrhizobium japonicum sips gene, homologous to prokaryotic signal peptidases, results in extensive changes in the expression of PBW-specific nodulins of infected soybean (Glycine max) cells		
JOURNAL	Mol. Microbiol. 18 (5), 831-840 (1995)		
MEDLINE	96422470		
PUBMED	8825087		
REFERENCE	2 (bases 1 to 31495)		
AUTHORS	Mueller,P.		
TITLE	Extended sequencing of a DNA fragment of B. japonicum adjacent to the cox operon		
JOURNAL	unpublished		
REFERENCE	3 (bases 1 to 31495)		
AUTHORS	Mueller,P.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-AUG-1995) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany		
REFERENCE	4 (bases 1 to 31495)		
AUTHORS	Mueller,P.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2000) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany		
REMARK	Sequence update by submitter		
COMMENT	On Jun 26, 2000 this sequence version replaced gi:1200339.		
FEATURES	Location/Qualifiers		
source	1. 31495		
	/organism="Bradyrhizobium japonicum"		
	/strain="USDA1105pc4"		
	/db_xref="taxon:375"		
	/clone="C27"		
gene	complement(7..909)		
	/gene="ephB"		
CDS	complement(7..909)		
	/gene="ephB"		
	/codon_start=1		

[illegible]



OY	1387	cacttcccttggaagcgtlcaagtgatgcgacggccgctttacatacgactatcccgaagccttcg	1446
Dd	23234	CACCTTCCTGGGCGCTGTCCGGGCATGCCGCCCGCTACGTAGACTATCCGACGCGTTCGCC	23175
OY	1447	ctgtggaacaagaatcctgcctctaagtgtacctgcgcttcgctcgcttcgtcttc	1506
Dd	23174	GSGCTGAACCTGTGCTCGTGCGGCTCCTTAATTTCGGGCTTGCGGCTCCGATCTTC	23115
OY	1507	accgtagctcttgtctatacgcctggttgcctgcgcgcgcgagaaaccgctcgaaccgtg	1566
Dd	23114	CTCTAATTGCGGTATGATGATCC---CTTTGGAAAAAAAGGTGCCCGCGTAGCAATCCGTTGG	23058
OY	1567	ggcgaattgcgcgatcacgaatgtaaggacgcctgcataccacgccttcgccgaacgcttc	1626
Dd	23057	GGTGGCGGGCGGACACACGCTGAGTGGACGCTGCGCTTCGCGCGCGCTTCATCATGACTTC	22998
OY	1627	gaaacgctgccc	1638
Dd	22997	GAAGTGTGCGCC	22986
RESULT	14		
R COXA			
LOCUS	R COXA	1707 bp	DNA
DEFINITION	R. leguminosarum coxa gene.		BCT
VERSION	X74341		25-APR-1994
KEYWORDS	X74341.1 GI:396698		
SOURCE	cytochrome c oxidase; cytochrome c oxidase subunit I.		
ORGANISM	Rhizobium leguminosarum.		
	Rhizobium leguminosarum.		
	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
	Rhizobiaceae; Rhizobium.		
REFERENCE	1 (bases 1 to 1707)		
AUTHORS	Gabel,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUL-1993) C. Gabel, The Johns Hopkins University,		
REFERENCE	Dept. of Biology, Charles and 34th Sts., Baltimore MD 21218, USA		
AUTHORS	2 (bases 1 to 1707)		
TITLE	Gabel,C.; Bittlinger,M.A. and Maier,R.J.		
	Cytochrome aad gene regulation in members of the family		
	Rhizobiaceae: comparison of copper and oxygen effects in		
	Bradyrhizobium japonicum and Rhizobium tropici		
	Appl. Environ. Microbiol. 60 (1), 141-148 (1994)		
JOURNAL	94161496		
MEDLINE			
FEATURES			
source	Location/Qualifiers		
	1..1707		
	/organism="Rhizobium leguminosarum"		
	/strain="USDA 27"		
	/db_xref="taxon:384"		
gene	91..1707		
	/gene="coxa"		
	91..1707		
	/gene="coxa"		
	/EC_number="1.9.3.1"		
	/codon_start=1		
	/transl_table=11		
	/product="Cytochrome-c oxidase, subunit I"		
	/protein_id="CAA52388.1"		
	/db_xref="GI:396698"		
	/db_xref="SWISS-PROT:Q0885"		
	/translation="MATSAAAHGEHAEDHGDHEAHPTGMRRSTNHHKDIGLYLIPAI		
	IAGVIGAMSLAIRAEIMYGVEYEENHTHLYNVFETSHGVIMIFPMVPAMIGGFNMW		
	FLPLMTGADMAPPRMNISFLLIRPASFCILLMETPEEGEAGACAGATMYPLSS		
	SGHPGAVDIAIPSILITAGASSILGALINFTTILINKRAPGTLRKMPLEFAVSYLITAF		
	LLESLPDLVLAGATITMLLDTRNFQTTFEPAREGGDPLLVOHLEFFGHREYVLTILPFG		
	GMTSHIISTFSRKPFVGVIQNVYMAAIGGLFYVMVAHYHIVGMDLEDTEAYEVSATIM		
	IIAAPTGIKIIFSWATMGGSIEPATLMALAEILEFTVGVTGVLANASIDRVLAH		
	DIYYVVAHFHVLSIGALIFAIFACMYTPMKMSGYMNELTAEAHFWLIIFGWLIFFE		
	PEHIGTISGMPPRYIDYPDPAFAGNIYSSISGYISGRSVLLFITCYVDAPRAKNYPVGD		
	NPMAGATITLEMILTSPRPHEEVRIVE"		
BASE COUNT	277 a	537 c	479 g
ORIGIN			414 t

[illegible]



PTFARKPLFGYTSWVYATASIAFLSFVVAHMHMFVGLPTVIGELFMVATMLIAVPLG  
 VKVFNWVTTWMEGSLTFEETPMLFAVAFLVITFTGGFSGIMLAIAPAFQVHDYFVA  
 HFHYLVPGATFGIPASATYVLKPTWGHMDETIGKLFHMKSFTGMLVAFPMHFVGL  
 AGMRRLPDVNLQFADFNWSSIGAFMGITQLLFLFTVICKINGKPADAKPMDGAE  
 GLEWSPSPAYHFSFPEVKR"

gene  
 /gene="PA0107"  
 6463..7017  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAG03497.1"  
 /db\_xref="GI:9945934"

CDS  
 /translation="MSDAKVDTRLVGRLLLVYVLMRSGFALVPLDYWCRLGNG  
 KITASATSGEODVQKREYVQETSTNNIMWFEFRAGOLVHPGAVQMYFAN  
 PSDKPMYQAIPISIAPEAAAYFHKTCFCFTQDVLQPGESIEPWFIVDRDLPKD  
 RHYVLAYLFDITARKPPVAGR"

gene  
 /gene="coi11"  
 /note="PA0108"  
 7028..7915  
 /gene="coi11"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome c oxidase, subunit III"  
 /protein\_id="AAG03498.1"  
 /db\_xref="GI:9945935"

CDS  
 /translation="MASHEHYVYPAOSKWPITASIGLTVTFGLTWENDLQAGHES  
 HGPVIFVQGLIAVLMFGVNYIRSRAGLYSAQMDRSFRGMSFTIFSEVWFPA  
 FPGALFYVRHFGAGLWEGGAGVAAHMLNFOYSWLLQTPDKLEPPSAVIEWP  
 LPLNTLLVTSFTVFAHNAALKKRGKPLKMLATVLTGLTAPLLOAEVVAVN  
 ELGLTLAGIYGSFEMLTGPHGAVTLGLLIIGIMLIRLRGHFPAEHFGEAASW  
 YMHVVDVWVIGLFFVYV"

gene  
 complement(7931..8140)  
 /gene="PA0109"  
 complement(7931..8140)  
 /gene="PA0109"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG03499.1"  
 /db\_xref="GI:9945936"

CDS  
 /translation="MLKVALVLLLATLVLSFGLFELVKDQHGSRVNSLTVRVL  
 AAATLVVAMGFSGELNSHAPMHF"

gene  
 8156..8950  
 /gene="PA0110"  
 8156..8950  
 /gene="PA0110"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG03500.1"  
 /db\_xref="GI:9945937"

CDS  
 /translation="MGCFMOYSQRLGEGANDRRSMKGAERPGLLFTLVVLLPVLL  
 WLCTQORADEKRALIASYERARGAEPVSGOLEGRPAVYRVLRHGFEDRHLL  
 LDNRNGAGVAVLQPFYDQASGLMLVLRGVAWTDRRSPITLFTPDVLLDWT  
 YLRPGLHLADAPAGMPRLVTLQDLPALMOGFRAGALPWEIRLLEPGDASFTDMP  
 VSMPEHHTGYAVQWALATALLATLVLYLGVRRARKNHESRSDA"

gene  
 8925..9503  
 /gene="PA0111"  
 8925..9503  
 /gene="PA0111"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG03501.1"  
 /db\_xref="GI:9945938"

CDS  
 /translation="MSLAIPIPERPRRARGRLQILAIGLVVGPMLLASAMYKNFV  
 PGRSYSGALIGNGOTPADLGVGRSGEOMLTVAPGCGECCOOLVYLAQVITIA  
 LGREASRAGHALAAGELPADFANLVRODPRIORGGRDPAIYDKRGGGERPLWYD  
 PHGNVLRYPARKNGKLLKDVOGLLKLSHIG"

gene  
 9568..10641

Query Match 30.8%; Score 515.6; DB 1; Length 12829;  
 Best Local Similarity 63.0%; Pred. No. 3.2e-98;  
 Matches 854; Conservative 0; Mismatches 484; Indels 18; Gaps 3;

CDS  
 /gene="PA0112"  
 9568..10641  
 /gene="PA0112"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG03502.1"  
 /db\_xref="GI:9945939"

QY 286 ctgatagtgtctcttgaggatgcccgatctgtcggtgttcttgtaactatctgatg 345  
 Db 5106 CTGATCATTGGTCTTGGCGCGGTATGCCGCTTGTGCGCCCACTGGATGATC 5165

QY 346 ccgctgcaaatcgcgctccggaatgagcctcccgatgacacacatgctgtcttg 405  
 Db 5166 CCGGTGATGATTCGGCGCGGACATGCGCTCCCGGATGACAACTTCACCTTCTG 5225

QY 406 ctgttcattgcggtacacgcgatggcggtgttcgtctgtcgacccggcggtgacg 465  
 Db 5226 CTCTGCGCGCGCGCTTGGCGCTGTGTCAGCACCTTGTTCATGCCCGCGG 5278

QY 466 cagctgggttcggcggtgtgtgtctgtgacccgcgctgcacccgggaagctgac 525  
 Db 5279 -----CGGCCCAACTTGGCTGACCTTATGCCGCGCTGACCACTTGGCCCG 5333

QY 526 taltcgaatgacctcggaatttcgagttcactgtgcgggttcctcctcgaatcattgac 585  
 Db 5334 CACAGCCTGACCTTTCATCTTCGACATCCACCTGACCGGATCAGTGCATGAGG 5393

QY 586 gcatcaacatgcatcagacctcttgacacatgacggcccgagatgacgtgcacaa 645  
 Db 5394 GCGATCAGCTGATCGGCACCATCTCAACCTTGGCGCGCGGATGACCCCTGATGA 6453

QY 646 gtgcgctgtctcgtgctgcatcattacacgctgtgacctgctcgtgagcgtgcg 705  
 Db 5454 ATGCCGCTGTCCTGTCGACCTGCTGATGATCAGCGCTTCTTTGATCCGGGATG 5513

QY 706 gtcttgagctgtgcaatcacaatgctgtgcacgcgaaacttc 765  
 Db 5514 GTCTGCGCGCGCTGTCGACATGATGATGATGATCATTGCGCACAGCTTCTTC 5573

QY 766 aatcctgctgtgagcggtacacgcatctctgtacacacacatcctgtgtcttgggac 825  
 Db 5574 AGCGCGCGCGCGCGCGCGACCGGCTGCTGTCCAGACGATGCTTGGTCTTGGCG 5633

QY 826 ccggaagtgtacatcattctgcgccgctgtgcatcatcagccatgtcgtgtgac 885  
 Db 5634 CCGAGAGTGATCATGATCTGCGCGCTTGGCGGTACAGTCCATTCGCCGAC 5693

QY 886 ttctcgaaaagcggctctcgttaccctgcgcgagtgtctatgcaatgtgtgcaatcgt 945  
 Db 5694 TWCGCCGCGCAAGCGCTGTTCCGCTACACCTCGATGCTACGCCACCGCCAGCAT 5753

QY 946 gtcttgagctgtctgctggcgacacacatgtaacacgtgtgtgtgtcgtgtgacg 1005  
 Db 5754 TTCCTCTCTTCTGCTGCTGTGCGCGCACACATTTGTGTGTGGCTTCCCGGTAC 5813

QY 1006 caatcctacttcatgtctgcacacatgtgatacggtgcgacgcgacatlaagatctc 1065  
 Db 5814 GAAGCTGTTCTTCAVGTACGCCACCATGTCGATGCGCGTCCCGCGGGTGAAGGTT 5873

QY 1066 tcttgatgcgacgctgtggtggcgctgtgtgattcaatcgccgactgtcgtggac 1125  
 Db 5874 AACTGGGTACCAACCATGTGGAGGGGTGCTGACACTTGGAGACCGCGATGCTTGG 5933























Db 7002 ctacacgacagctactcgttctagcccaactccactccatctgctctatcaatagagctgta 7061  
 QY 1261 ttgagatcttcggtctactcttctacatcggaagttctcggtcggtctccg 1320  
 Db 7062 ttgagatcttcggtctactcttctacatcggaagttctcggtcggtctccg 1320  
 QY 1321 gaatggctgcaagaagctctctgacatcttctcattcggtgcaagctcgtctc 1380  
 Db 7122 caaacctacgcaaaatccattccattccatccatccatccatccatccatccatcc 7181  
 QY 1381 ccgagacactctctggaagctggaatgagcgccgctgtaacatccgaagcc 1440  
 Db 7182 ccacacactctctcgtctccatccggaatgcccgcagctgtaacatccgaagcc 7241  
 QY 1441 ttgagctggaagaagctctgctcgtggtggtctggtctggtctggtctggt 1500  
 Db 7242 tacacacagaaacatccatccatctgtagctcattctctctctctctctctct 7301  
 QY 1501 ttcttcacgtgactctt 1518  
 Db 7302 ttaataattctgatt 7319

## RESULT 8

AAZ57203 standard; DNA; 16569 BP.

AAZ57203:

30-MAR-2000 (first entry)

Human mitochondrial DNA sequence SEQ ID NO:2.

Human: mitochondrial DNA; extramitochondrial DNA; mtDNA; exmtDNA;  
 diagnosis: quantification; detection; dystonia; Alzheimer's disease;  
 Huntington's disease; Parkinson's disease; schizophrenia; stroke;  
 non-insulin dependent diabetes mellitus; mitochondrial encephalopathy;  
 lactic acidosis; myoclonic epilepsy ragged red fibre syndrome;  
 Leber's hereditary optic neuropathy; ds.

Homo sapiens.

WO966075-A2.

23-DEC-1999.

14-JUN-1999: 99MO-US13426.

15-JUN-1998: 98US-0097889.

15-JUN-1998: 98US-0098079.

30-APR-1999: 99US-0302681.

(MITO-) MITOKOR.

Herrnstadt C, Ghosh SS, Cleverger W, Fahy ED, Davis RE:

Quantification of extramitochondrial DNA for diagnosis of, e.g.

Alzheimer's, Huntington's and Parkinson's disease -

Claim 29; Fig 2; 157pp; English.

The present invention describes a method for the quantification of  
 extramitochondrial DNA (exmtDNA) by determining the ratio of a first  
 and second biological sample containing exmtDNA and mitochondrial DNA  
 (mtDNA) to determine the risk or presence of a disease associated with  
 altered mitochondrial function. The method can be used to determine  
 the risk of or presence of a disease associated with altered  
 mitochondrial function, especially Alzheimer's disease, Huntington's  
 disease, Parkinson's disease, dystonia, schizophrenia, non-insulin  
 dependent diabetes mellitus, mitochondrial encephalopathy, lactic  
 acidosis, stroke, myoclonic epilepsy ragged red fibre syndrome and

CC Leber's hereditary optic neuropathy. The method can also be used to  
 CC identify agents suitable for treating such diseases, in particular  
 CC Alzheimer's disease, AAZ57202 to AAZ57313 represent nucleotide sequences  
 CC used in the exemplification of the present invention. More specifically  
 CC AAZ57206 to AAZ57313 are PCR primers used in the detection of exmtDNA  
 CC and mtDNA.

Sequence 16569 BP; 5122 A; 5180 C; 2171 G; 4096 T; 0 other;

Query Match 23.7%; Score 397.2; DB 21; Length 16569;  
 Best Local Similarity 58.1%; Pred. No. 3.9e-96;  
 Matches 743; Conservative 0; Mismatches 52; Indels 12; Gaps 2;

QY 244 aacgacacccctggaagctcgtctacatccatccatccatccatccatccatccatcc 303  
 Db 6051 aacgacacacatcaaaagctatcgtctacatccatccatccatccatccatccatcc 6110  
 QY 304 ggtatcccgcatctgtctggtctgtgtaactatctgacgtcgctgcaaatccgagct 363  
 Db 6111 gtaataccatcataatccgaggtcttgcaactgactagttcccatataatccgtgccc 6170  
 QY 364 ccggaatagccttcccgatgataaacaactgtctgtctgctgctgctgctgctgctg 423  
 Db 6171 ccggaatagccttcccgatgataaacaactgtctgtctgctgctgctgctgctgctg 6230  
 QY 424 gcgaatggcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 483  
 Db 6231 ctactccgtctgcatctgctgctgctgctgctgctgctgctgctgctgctgctgctg 6287  
 QY 484 ggttggtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 543  
 Db 6288 -----taaccctcccttgacgaggaactcctccacccctgagctcctgagctaac 6341  
 QY 544 attctcggtcactctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 603  
 Db 6342 attctccttaccactgacgagtgctcctcctatctttagggccatcaattcacaaca 6401  
 QY 604 accctcttgaacatgacgaccccgacatgaacgtcgaacgaatgacgtctgctgctg 663  
 Db 6402 acaattatcaataaaacccctgcataaccacatacaacgaacccctctgctgctg 6461  
 QY 664 tctgactcttatacagctgctgctgctgctgctgctgctgctgctgctgctgctgctg 723  
 Db 6462 tccgtccttaatacagcagctccttctcctatctcctccagctcctgctgctgctg 6521  
 QY 724 accatgctgctgacccgactgactcctgacgacgactccttcaatccgtcgtggtggt 783  
 Db 6522 actataactatacagacccgacactcaacacacactccttcgaccccggtggtggtg 6581  
 QY 784 gaccgactctgtacaaacacatcctgtgtctcttctggtgacccggaagtgtacatc 843  
 Db 6582 gaccacatctatacaacaacactctcgtatttctgctgacccctggaagtgtacatc 6641  
 QY 844 attctgcccgtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 900  
 Db 6642 atcctacacagctctggaataatcccatatgttaacttactactccggaataaaagaa 6701  
 QY 901 gctctggttactgocagatgctgctgctgctgctgctgctgctgctgctgctgctg 960  
 Db 6702 ccaattgataataatgataatgctgctgctgctgctgctgctgctgctgctgctgctg 6761  
 QY 961 gcttggtgacacacatgctgctgctgctgctgctgctgctgctgctgctgctgctg 1020  
 Db 6762 ggttgagacacacacataatcttcaagtagaagcgtgagacacacacacacacac 6821  
 QY 1021 ctggacacacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1080  
 Db 6822 tccgtacacataatcgtctgctgctgctgctgctgctgctgctgctgctgctgctg 6881  
 QY 1081 atgttggtggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1140  
 Db 6882 ctcaacggaagacataatgaaatgactctgctgagtgctgctgctgctgctgctgct 6941











Db 812 cgtcttcggcattgtcttcgagatcattcctgtgtcttcctccgtaagcaatgttcgatt 871  
 QY 911 acctgcgaatggtctctgcaatggtggcaatcgtgtcttcgtggccttgcgttcggcgc 970  
 Db 872 agctgcgcgtgattcttcgcaactgttcgaattgttgcaatcgaatggtctgtggcctc 931  
 QY 971 accacatgtaacacgttggatgttcgtaccacgaatcctactcatgctcgtcggcaca 1030  
 Db 932 accaatgttcttactgtgcgcagtttgccttccttccttccttcacg-----agct 985  
 QY 1031 tgggttcggtgtgcgcgaacgcatgaatctctcgttgatcgcgaatgttgggctc 1090  
 Db 986 tctgtattcgtcttcacgcggttaagtcttcgaactggtgtgaacacatgtggaag 1045  
 QY 1091 gctcgttgaatgcaatcgcgcgaatcgttcgttcgttcgttcgttcgttcgttcacg 1150  
 Db 1046 gtaactcaacttgggaaccccaatcgtcgttcgttcgttcgttcgttcgttcgttc 1105  
 QY 1151 tgggtgtgtgacgcggtatcgtgtgcgcgaacggttcgttcgttcgttcgttcacg 1210  
 Db 1106 tctgtgttcgttcgacgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1165  
 QY 1211 cctattacgt 1270  
 Db 1166 cctactcctgc 1225  
 QY 1271 tgcgcgttactacttcttcaatgcgaatgttcgcgcgcgcgcgcgcgcgcgcgcgcgc 1330  
 Db 1226 gtcgagcggttacttctgttccggaatgactgtgcgcgcgcgcgcgcgcgcgcgcgc 1285  
 QY 1331 caaagctgacttctgcgaacttcttcgttcgttcgttcgttcgttcgttcgttcgc 1390  
 Db 1286 gaaagctcacttctgttcgaacttcttcgttcgttcgttcgttcgttcgttcgttc 1345  
 QY 1391 tctgtgagcgtcgaatgttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 1444  
 Db 1346 ggtgtggaacatgtggtatgcaacgttcgaacgttcgaacgttcgaacgttcgaac 1405  
 QY 1445 cgtgtggaacaaatgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 1504  
 Db 1406 ccatcacaacagatctccacagctgttccttccttcgttcgttcgttcgttcgttc 1465  
 QY 1505 tcatgtgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1564  
 Db 1466 tcatctggaacgttcttcaagcttcgttcgttcgttcgttcgttcgttcgttcgt 1525  
 QY 1565 ggggcgaatcgcgcgaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 1624  
 Db 1526 gggg---ttacgcaactcctcgtggaagtgggaacactcctcctcctcctcctc 1582  
 QY 1625 tcgaaacgctgcc 1637  
 Db 1583 tcgcatccttgcc 1595

RESULT 13  
 AAH68532/c  
 ID AAH68532 standard: DNA; 349980 BP.

AC AAH68532;  
 XX 26-SEP-2001 (first entry)  
 DE C glutamicum coding sequence fragment SEQ ID NO: 7067.  
 XX  
 KW Corynebacterium: amino acid synthesis; vitamin; saccharide;  
 XX organic acid synthesis; ds.  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX

PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KIOW ) KYOMA HAKKO KOYO KK.  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 DR WPI: 2001-376931/40.  
 XX  
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT

Disclosure: SEQ ID NO: 7067; 246pp + Sequence Listing: English.

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from corynebacterium bacterium, and identifying a homologue of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 349980 BP: 80900 A; 98397 C; 92139 G; 78544 T; 0 other;

Query Match 17.7%; Score 296.6; DB 22; Length 349980;

Best Local Similarity 54.0%; Pred. No. 1e-68; Matches 752; Conservative 0; Mismatches 614; Indels 27; Gaps 6;

QY 251 acctgtggaacatgattgttaccatcgtatcgttcgttcgttcgttcgttcgttcgttc 310  
 Db 272582 AGCAATTGAACACACTGTTCACATGCAACGAACTGTGATCTCTCTGTACGAA--- 272524  
 QY 311 ccgaatgttctgtgttttggtaactatcgtatgcgcgttcgaatgcgcgcgttcgata 370  
 Db 272525 CTCGAATTGTTGGGGTTCCTTAACGTCGCACTTCAAGATCGGCGCTGACG 272466  
 QY 371 tggcttcgcgcgtatgaacactgttcgttcgttcgttcgttcgttcgttcgttcgttc 430  
 Db 272465 TAGCTTCCACAGTTTAATGCTTCGCTTCGATGACACCGGTGGGTGCGCA 272406  
 QY 431 gcgtgctcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgt 490  
 Db 272405 TGCTGACCGGCTTCGACCCCGGGGTGTCTGCCGACTTCGTTGGACCATGTACTCC 272346  
 QY 491 ttcgtacccgcgcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 550  
 Db 272345 CACGTGCTGACGAAATTCACCTCCCAAGCCTTGCTCT-----GACATGTGATGTGCG 272292  
 QY 551 cgttcacttgcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 610  
 Db 272291 GTGTCGTCGCAACTGATTTGGCTGCTTCGCAATTAACATGCTCACCAATCC 272232  
 QY 611 tgaacatgc 670  
 Db 272231 TCTGCTCCCGCGCACCTGTATGACATGTTGCGTATGCTATTTTACCTGGAATATCT 272172  
 QY 671 ttatcacggttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 730

[illegible]

RESULT	14
AAAF71686	
ID	AAAF71686 standard; DNA; 1779 BP.
XX	
AC	AAAF71686;
XX	

XX	30-APR-2001	(first entry)	
DE	Corynebacterium glutamicum SMP protein nucleotide sequence	SEQ ID NO:653.	
XX	Corynebacterium glutamicum; carbon metabolism and energy production;		
KW	SMP protein; sugar metabolism and oxidative phosphorylation protein;		
KW	fine chemical production; organic acid; proteinogenic amino acid;		
KW	nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;		
KW	nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;		
KW	carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;		
KW	diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.		
XX			
OS	Corynebacterium glutamicum.		
PN	WO200100844-A2.		
XX			
PD	04-JAN-2001.		
XX			
PF	23-JUN-2000; 2000WO-IB00943.		
XX			
PR	25-JUN-1999; 99US-0141031.		
PR	08-JUL-1999; 99DE-1031412.		
PR	08-JUL-1999; 99DE-1031413.		
PR	08-JUL-1999; 99DE-1031419.		
PR	08-JUL-1999; 99DE-1031420.		
PR	08-JUL-1999; 99DE-1031424.		
PR	08-JUL-1999; 99DE-1031428.		
PR	08-JUL-1999; 99DE-1031431.		
PR	08-JUL-1999; 99DE-1031433.		
PR	08-JUL-1999; 99DE-1031434.		
PR	08-JUL-1999; 99DE-1031510.		
PR	08-JUL-1999; 99DE-1031562.		
PR	08-JUL-1999; 99DE-1031634.		
PR	09-JUL-1999; 99DE-1032180.		
PR	09-JUL-1999; 99DE-1032227.		
PR	09-JUL-1999; 99US-1032230.		
PR	09-JUL-1999; 99US-1032208.		
PR	14-JUL-1999; 99DE-1032924.		
PR	14-JUL-1999; 99DE-1032973.		
PR	14-JUL-1999; 99DE-1033005.		
PR	27-AUG-1999; 99DE-1040765.		
PR	31-AUG-1999; 99US-0151572.		
PR	03-SEP-1999; 99DE-1042076.		
PR	03-SEP-1999; 99DE-1042079.		
PR	03-SEP-1999; 99DE-1042086.		
PR	03-SEP-1999; 99DE-1042087.		
PR	03-SEP-1999; 99DE-1042088.		
PR	03-SEP-1999; 99DE-1042095.		
PR	03-SEP-1999; 99DE-1042123.		
XX	03-SEP-1999; 99DE-1042125.		
RA	(BADI ) BASF AG.		
XX			
PI	Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habernauer G;		
XX			
DR	WPI; 2001-061975/07.		
XX	P-PSDB; AAB79569.		
PT	New isolated Corynebacterium glutamicum nucleic acid encoding a sugar		
PT	metabolism and oxidative phosphorylation protein for production or		
PT	modulation of production of fine chemicals e.g. amino acids,		
XX	carbohydrates or enzymes -		
PS	Claim 3; Page 1069-1072; 1246pp; English.		
XX			
CC	AAAF1360 to AAFF1750 encode the Corynebacterium glutamicum sugar		
CC	metabolism and oxidative phosphorylation (SMP) proteins given in		
CC	AAAB79244 to AAB 79633 which are involved in carbon metabolism and		
CC	energy production. The C. glutamicum SMP gene can be used in vectors		
CC	(II) for expression in host cells and production or modulation of		
CC	production of fine chemicals, such as, an organic acid, a proteinogenic		
CC	or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,		
CC	a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty		





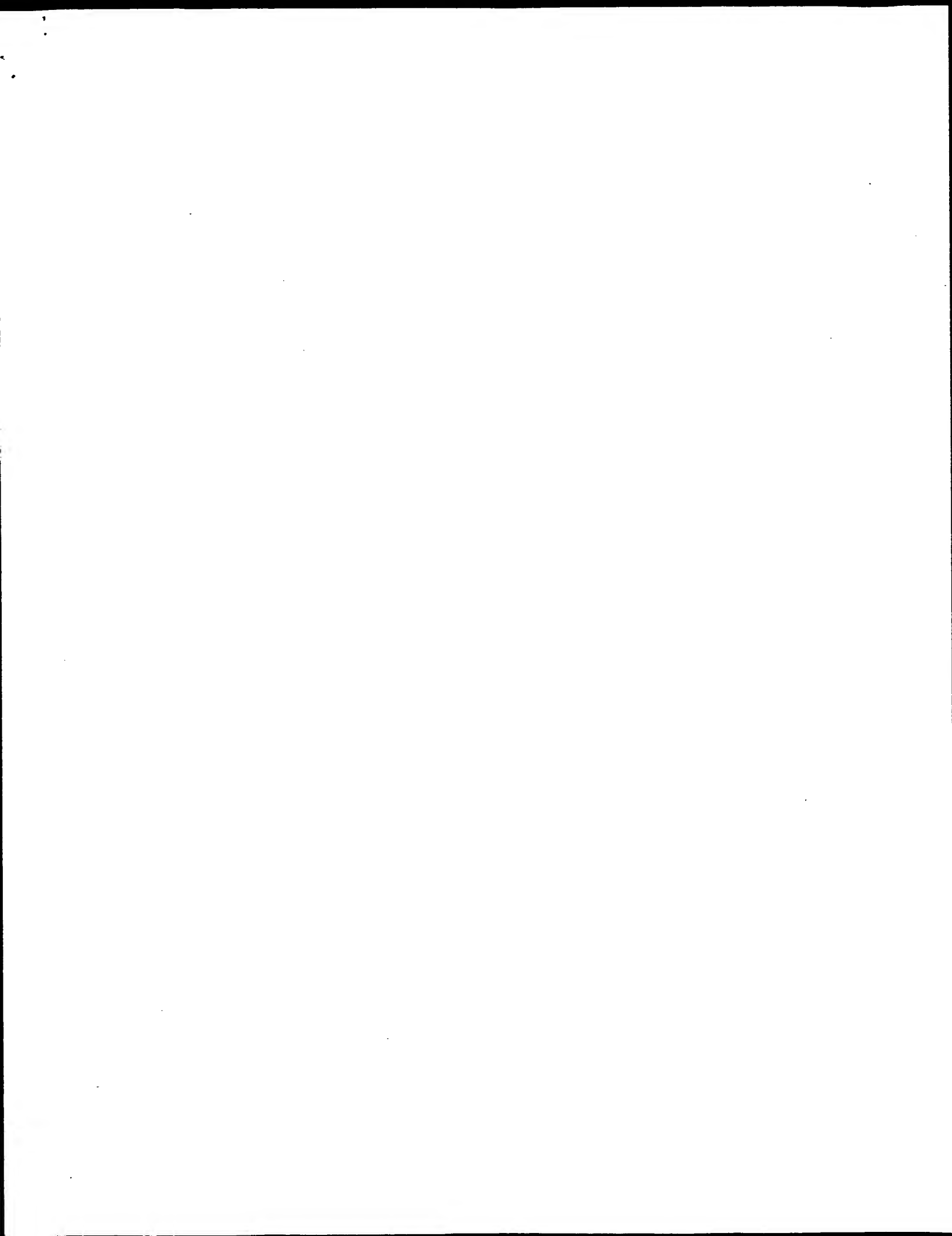




Thu Dec 27 08:21:16 2001

us-09-712-768-1.rng

---





QY	304	ggaatcccgcatctgtctgggtgtttgttgtaactctatcgatgacgctgcaaatcgagcgt	363
Db	279	GTAAATACCCTCAATCAATCGAGGCGTTTGGCAACTACTAGTTCCCTCAATATATGGTGGC	338
QY	364	ccgataatgacgttcccgcgatgaaacaacactgcttcgtctgttccatgtccggtaac	423
Db	339	CCGATATGGCGTTTCCCGCATTAATACACATATAGCTTCTGACTCTTAACCTCCCTCTC	398
QY	424	ggcattggcgctgcgttcgctgtctgttcgcaacggcggtgaacgttcagctgtggttcggcgt	483
Db	399	CTACTCTCTGCTCCACTCTGCTATATAGGAGGCGGAGGAGCAAGTGTGAACAGTCT--	455
QY	484	gattgggtctgtacaacggcgctgtcgaccgcggaagctggcgtatctgattgacctcgcg	543
Db	456	-----TACCCTCCCTTAGAGGAGGAACTACTCCCAACCTCGAGGCGCTCCGTAGACCTAAC	509
QY	544	atttcgggttcaactgtgtcgggtgcctccctcgataatggtgcgatacaatgatacgt	603
Db	510	ATCTTCTCTTACACCTAGCAGAGGTGTCTCTCTACTTATAGGGGCGCATCAATTTCTACA	569
QY	604	acctctctgaacctgacggcgcccgccgcaatgaaacgtgacgaatgacgtgtctcgtgg	663
Db	570	ACAATTATCAATATTAACCCCTCGCATATACCCAATACCAACGCCCTCTTGCTGTGA	629
QY	664	tgcatacttatacagctgttgctgtatcctcgtcgtgcgtgcggtctcgtgtgtgcaat	723
Db	630	TCCCTCTTAATACAGCAGAGCTACTCTTCTCTATCTCTCCAGTCTCTAGCTGTGGATC	689
QY	724	acgatgctgtcgaacggcgcttaacttcggcagacacttcaatctcgtcgtggcggt	783
Db	690	ACTATTACTACTAACAGCCCAACCTCAACACCACTTCTTGACACCCCGCAGAGAGGA	749
QY	784	gaccagatctctgtaaccaacaacatcctgtgttcttccttgggacccggaaagtatacatc	843
Db	750	GACCCATTTCTATACCAACACTATTTCTATTTTTCGGTCACTCGAGTTTAAATCTTT	809
QY	844	attctgcggcgtttgcatcatcacaacgactgtcgtgtgacacttctc---gaaaaagcg	900
Db	810	ATCTCACCAGGCTTGGAAATATCTCCCATATGTATACTTACTCTCCGGAATAAAGAA	869
QY	901	gtcttcggtttacactcgcgaatggtcctatgcaatgtggtgcaatcgtgtctgtgactttgc	960
Db	870	CCATTTGGATATCAATAGGTATGTCGTAGGCTATGATATCAATTTGAGTTCTTAGGGTTTATC	929
QY	961	gtctggcgcgcaacaatgataccgctgtgtatgtcgtgtgacccagcaatcctactatgt	1020
Db	930	GTGTAGAGCACTCAATATATTTTACAGTATGGAATATGACGTATGACACACAGCATATTTCAAC	989
QY	1021	ctgggcacaacatggtgatacgcgtgtgcgcgaacccgagcatlaaatcttcctcgtgatacgcga	1080
Db	990	TCCGCTACCAATATCATTCGCTATTCGCCACCGGGGTCAAGTATTTAGCTGACTCGCCACA	1049
QY	1081	atgttgaggcggtcgtgttgatgttcaaatctgcgaatgctctcgttggtcctttggtttatgtc	1140
Db	1050	CTCCACGGAACCAATATATGAATATGATCTGCTGCAAGTCTCTGAGCCCTTATGAGATTAATCTTT	1109
QY	1141	ctgtttacacgtgggtgtgttgaaacggtatcgtctgtgcaacggcggtctgtgaacggtga	1200
Db	1110	CTTTTACACCGTATGAGTGGCTGTACTGGCATTTGTATAGCAACTCATCATATAGCAATCTGTA	1169
QY	1201	tatacgaacacttatacgtgttggtgcgacttccattcatatgtatgtcgtctgggtgacatc	1260
Db	1170	CTACACGACAGCTACTGCTGTATGGCCCACTTCCATATGTCTCTATACATATAGCACTGTA	1229
QY	1261	tttgcgacttcgcgcggtatctactctttaaagtcgaagttctcgtggcgcgcttcccg	1320
Db	1230	TTTTGCATCAATAGAGGCGTTCATTACTGATTTTCCCTATTTCTCAAGGCTACACCTTAGAC	1289
QY	1321	gaatggcgctgaagctgtcaactctgtgaacttctcatcgtgtgcgaacttcaactgttc	1380
Db	1290	CANAACCTACGCCAANAATTCATTTCACTATCATATTCATCTATGCGGGTAATATCTTAACCTTTCTTC	1349
QY	1381	ccgacgaacttctcgtggagctcaggtatagtcgcgcggtttacatcgaactatccgaaagcc	1440

[illegible]

RESULT 2  
US-08-451-096-1

; Sequence 1, Application US/08451096  
; Patent No. 5760205

```

; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.

```

APPLICANT: HerinStadt, Corinna  
TITLE OF INVENTION: Diagnostic  
; ;

```
; TITLE OF INVENTION:  FOR ALZHEIMER  
; NUMBER OF SEQUENCES: 95  
; ADDRESS/DOWNLOAD ADDRESS:
```

;; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flore  
; CEMEM 4370 13 10113 willac

STREET: 4370 La Bolla Village  
CITY: San Diego  
STATE: California

STATE: CALIFORNIA  
COUNTRY: USA  
ZTB: 92122

```

;      ZIF. 02122
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: F10000 disk

```

; MEDIUM TYPE: FLOPPY DISK  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;   OPERATING SYSTEM:  PC DOS/MS
;
;   SOFTWARE:  PatentIn Release #1
;
;   CURRENT APPLICATION DATA:

```

APPLCATION NUMBER: US/08/451  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLCATION NUMBER: US 08/219  
FILING DATE: 30-MAR-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-AQ

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949  
INFORMATION FOR SEO ID NO: 1

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
;

```

```

;      TOPOLOGY:  linear
US-08-451-096-1

```

Query Match	23.7%;	Score
Best Local Similarity	58.1%;	Pre

Matches 743; Conservative 0;

QY 244 aacgacacctgtggaacgtcatgtt  
|||||

Db 219 AACGACCACATCTACAACGTTATCGTC

QY 304 ggtatccccgcattgttcggtgtttt

Db 279 GTAATACCCATCATATCGGAGGCTT

QY 364 ccgatatgacctccgcgtatgaac  
||| ||||| ||| ||| ||| |||

Db 339 CCGATATGGCGTTCCCGCATTAAG

100

OY	424	ggagatggcgagcgctgcgttcgttccgaacccggcggtgcagctcagctgggtttcggcgctt	483
Db	399	cttactctcgtctgcattctgctatattgtggagggccggagccaggaacaggtttaacaaagtc---	455
OY	484	ggttggtctctgtacccgcgcgtctcgaaaccgcgaagctgcatactcgaaagaccctccgcg	543
Db	456	-----TACCCTCCCTTATGACAGGGAGACTACCCACCCTGGAGGCTCCGTACAGACTAAC	509
OY	544	atttcgcgggttcaacttgctgggtgtgcctctcgatactggcgcgatcaacaatgatcaacg	603
Db	510	attctctctttagacacttagcaggggtgtcttccatcttttagggggccatcaattttcatcaca	569
OY	604	accctctgaaacatgcgcgcgcgcgcgcgaatgcgtctgaaagtgcaagtcgctgtctcgttg	663
Db	570	ACAATTATTCATATATAAAACCCCGTGCATATACCAATACCAAAAGCCCGCTTCGTCTGA	629
OY	664	tcgaacttatacaagcgtcttgctgcataccctgcctgcgtcgcgttctgcgttgatgcac	723
Db	630	ttccgtcttaattacacagagatgcttacttctccatctctccacgtccagtcgtgcgtac	689
OY	724	accagctgcctgcagccgcgaacttgcgaagacacttctcaatctctgcgtgcgcgt	783
Db	690	ACTATACTACTAACAAGACCGCACTCAACACACCTTCTTCACACCCCGCGGAGGAGGA	749
OY	784	gaaccgatctctgaaccaacatctcgttgcttcttgggaaccgcgaagtgcatacaatc	843
Db	750	GACCCCATTTCTATACCAAAACCTATTCTGATTTTTCGGTCAACCTGAACTTAATTTCT	809
OY	844	attctgcgcgcgttggcatcatatgaacatgctcgtctgcacctctc---gaaaagccg	900
Db	810	ATTCCTACAGGCTTCGGAAATATATCTCCATATATGTATATCTACTACTCGGAAATAAGAA	869
OY	901	gtcttcggttaacctgcgaatggtctatgcgaatggtgcgaatgcgaatctcaactatc	960
Db	870	CCATTTGGATATCATAGGTATAGTGTGCTGAGTATGATATCATTTGGATTCCTAAGGTTAT	929
OY	961	gtctggcggaaccacatgtaacacctgttgatgtcgtctgaaccgaatctcaactatc	1021
Db	930	GTGTGAGCAGCACCTATATATTTTAAAGTAGAGAGCTGATACACAGACCATATATTTAC	989
OY	1021	ctggacacacatggtgatcgcgtgcgcgaaccgcgaatcaagatctctcgttgatgcga	1080
Db	990	TCCGCTACCATTAATCATTCGTATTCGCCACGGCGGTCAAGTATTATAGCTACGCGACA	1044
OY	1081	atgctggggcgctcgcgttgatgttcaaatcgcgaatgcctctgcgccttggtcttatgtc	1140
Db	1050	CTCCAGCGGAAGCATATATGAATATCATCTGCTGCAGTGCCTGAGCCCTTAGATTCATCTTT	1109
OY	1141	ctgttcaacctgggttgggtgcacggatcagtcgtgcgcgaagcggtgtctgaacctgtca	1200
Db	1110	CTTTTACCGGTAGTGGCTGTGACTGGCATTTGTATTAGCAAACTCATCACTTAACATTCGA	1165
OY	1201	tatcaacacacattacgtgtgtgcgcacatctcaatgatgatgtcgtctgggtgcagat	1260
Db	1170	CTACACACACAGTACTACGTTTGAAGCCCACTTCACATATGTCTATCATATAGAGACTGTA	1225
OY	1261	tttgcgatcttcgcgggtatcaactcttttaacatgcgaagttctgcggcgcgcttcccg	1320
Db	1230	TTTGGCATTCATAGAGAGGCTTCATTTCACTGATTTTCCCTATTTCTCAAGGCTACACCTTAG	1288
OY	1321	gaatggagctbcaagctgcgaactcttgacacttctcatcgtctgcgaagctgaactgtctc	1380
Db	1290	CAAACTTAGCCCAAAATTCATTTCACTATCATATTCATGCGGTAAATCTTAACTTTCTTC	1349
OY	1381	ccggaagacactctctgggaagctcaaggatgcgcgcgcgtttacatcagcatctccgaagc	1440
OY	1441	ttcgcgcgtgcggaagaagtctgcatactgtctgtctcctgcgccttcgcgtctcgt	1500
Db	1410	TAGACCAACATGAAGAACATTCATATCTGTAGGCTCATTTCTTCTTAACAGCACTAATA	1469
OY	1501	ttcttcaatcgtgatccttt	1518

```

DB      1470  TTAAATATTTCATCAT 1487
      11 11 11 11 11
RESULT  3
US-08-810-599-1
; Sequence 1, Application US/08810599
; Patent No. 5976798
; GENERAL INFORMATION:
; APPLICANT: PARKER, W. Davis
; APPLICANT: HERRNSTADT, Corinna
; APPLICANT: GHOSH, Soumitra S.
; APPLICANT: FAHY, Edin
; TITLE OF INVENTION: Methods for Detecting Mitochondrial Mutations
; TITLE OF INVENTION: Diagnostic for Alzheimer's Disease and Methods for Determin
; TITLE OF INVENTION: of Mitochondrial Nucleic Acid
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/810,599
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/757,438
; FILING DATE: 27 NO. 5976798 1996
; APPLICATION NUMBER: US 08/614,072
; FILING DATE: 12 Mar 1996
; APPLICATION NUMBER: US 08/536,036
; FILING DATE: 29 Sep 1995
; APPLICATION NUMBER: US 08/414,969
; FILING DATE: 31 Mar 1995
; APPLICATION NUMBER: US 08/413,740
; FILING DATE: 30 Mar 1995
; APPLICATION NUMBER: US 08/410,658
; FILING DATE: 24 MARCH 1995
; APPLICATION NUMBER: US 08/397,808
; FILING DATE: 3 Mar 1995
; APPLICATION NUMBER: US 08/219,842
; FILING DATE: 30 MARCH 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2105/17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-810-599-1
Query Match      23.7%; Score 397.2; DB 2; length 1613;
Best Local Similarity 58.1%; Pred. No. 7 2e-99;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

```

244 aacggacacgtgtagacgtatcctaccatgtagtattcgtatcgttctgtg 303  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 219 AACGACACATCTACACAGCTTATGCTCACAGCCCATGATTTGTAATATCTCTTCAATA 278  
 304 ggtatcccgcatctgtcgtgtgttctgtaatactatcgtatcgctgcaalegagct 363  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 279 GTAATACCATCATATATGAGAGGCTTTGGCACTGATAGTTCCCTTAATATCGGTCC 338  
 364 ccggtatgagccttcgagatgtagaacaacgtctgtctgctgttctatcgagctacc 423  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 339 CCCGATATGCGCTTCCCGCATTAACACATTAACCTTCTGACTTTACCTCCCTCTCTC 398  
 424 gcatagagcgctgtcgtctgtcgcacggcggtgaaagctcagctgtgtcgtgagct 483  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 399 CTACTCTCTGCTGCAATCTCTATAGTGAGAGCCCGGAGAGCAAGGTTGAACAGTCC 455  
 484 ggttggtctctaccgctgtcgcacggcgagagctgtgctatcgtatgagctgcg 543  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 456 -----TACCCTCCCTTACAGGAGAACTATCCACCCCTGGAGCCCTGATAGACCTTACC 509  
 544 atttcgagctcactgtcgtgtgtcctcctcgtatcgtatgagatcaatacgaatgacg 603  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 510 ATCTTCTCTTACACTTACAGTATGCTCTCTCTATCTTATAGGAGCCATCATTTTCAACA 569  
 604 accctctgaacatgagcgcccccgcatgacgctgcaaaagtgccgtgtgtcgtgag 663  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 570 ACAATATCAATATAAACCCCTGCAATACCAACGACCCCTCTTCTCTGCA 629  
 664 tcatcttatacaggtgtgtgtatcgtcgtcgtgagcggtgtgtgtgtgtgtgtgtgt 723  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 630 TCCGTCCTATACACAGCACTCTCTACTTCTCTATCTCTCTCTCTCTCTCTCTCTCTCT 689  
 724 accatgctctacacgacgtatcgtcgcacgacgtctcctcctcgtcgtgagcggt 783  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 690 ACTATACTACTAAGACAGCCGCAACCTCAACACACACCTTCTTGAGACCCCGGAGAGAGA 749  
 784 gaccgcatctgtacacacacatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 843  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 750 GACCCATCTATACACACACCTATCTGATTTTTCGGTACCCCTGAAGTTATTTCTT 809  
 844 atttcgagcggt 900  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 810 ATCTTACCAAGGCTTGGAAATATCTCTCTATCTTACTTACTCTCGGAAATAAGAA 869  
 901 gcttcggtatcgtcgt 960  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 870 CCATTTGATATACATAGTATGCTCTAGTATGATATCAATTTGATTTCTAGGTTATTC 929  
 961 gcttgagcgacacacatgtaacacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 930 GTGTGAGACACCATATATTTTACATAGTAAGAAATGACGTAGACACACGACATATTTCCACC 989  
 1021 ctggacacacatgtagtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 990 TCCGCTACCAATATATCTGATCTCCACACGCGTAAAGTATTTAGCTGACTCGCAACA 1049  
 1081 atgtgagcgcgctgt 1140  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1050 CTCACAGGAAAGAAATATGATATGCTGAGAGTGTCTGAGACCTTAGATTCATCTTT 1109  
 1141 ctgttaccgt 1200  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1110 CTATTACACCTAGTGTGCTGACCTGCTATTTGTATAGCAAACTCATCTAGTACATGTA 1169  
 1201 tatcagacacattatagcgt 1260  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1170 CTACACGACACGACTACGTTGAGCCCACTTCCACTATGCTCTATCAATAGAGCTGTA 1229  
 1261 ttctgacatctcgt 1320  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1230 TTTTGCCATCTATAGAGAGGCTTCACTACATGATTTCCCTATTTCTCAAGGCTACAGCCTAGAGC 1289  
 1321 gaatggtgtgcaaaagctgtacatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380

Db 1290 CAACCTACGCCAAATTCATTTCTACTATCATATTCATTCGCGCAATCTAACCTTCTTC 1349  
 1381 ccgagacatcttcctggagcgtcaggtatgctgcgccccgtgtacatccgaagcc 1440  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1350 CCACAAACCTTTCTCGGCTATCCGAAATCCCGACGTTACTGCACTACCCGATGCA 1409  
 1441 ttcgctgtgaaacaaagctcgtctcctatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1410 TACACCAATGAACAAATCTATCATCTGTAGGCTCATCTATTTCTCTAACAGCAGTAATA 1469  
 1501 ttcttcatcgtatcctt 1518  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 1470 TTAATATTTTCATGAT 1487

RESULT 4  
 US-09-097-889-2  
 : Sequence 2, Application US/09097889  
 : Patent No. 6218117  
 : GENERAL INFORMATION:  
 : APPLICANT: Herrnsdorf, Corina  
 : APPLICANT: Ghosh, Soumitra S.  
 : APPLICANT: Davis, Robert E.  
 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING  
 : TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE  
 : NUMBER OF SEQUENCES: 26  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: SPED and BERRY LLP  
 : STREET: 6300 Columbia Center, 701 Fifth Avenue  
 : CITY: Seattle  
 : STATE: Washington  
 : COUNTRY: USA  
 : ZIP: 98104  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/097,889  
 : FILING DATE: 15-JUN-1998  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Rosenman Ph.D., Stephen J.  
 : REGISTRATION NUMBER: 43,058  
 : REFERENCE/DOCKET NUMBER: 660088.417  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (206) 622-4800  
 : TELEFAX: (206) 682-6031  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 16569 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : US-09-097-889-2

Query Match 23.7%; Score 397.2; DB 4; Length 16569;  
 Best local similarity 58.1%; Pred. No. 1.9e-98;  
 Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

244 aacggacacgtgtagacgtatcctaccatgtagtattcgtatcgttctgtg 303  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 6051 AACGACACATCTACACAGCTTATGCTCACAGCCCATGATTTGTAATATCTCTTCAATA 6110  
 304 ggtatcccgcatctgtcgtgtgttctgtaatactatcgtatcgctgcaalegagct 363  
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
 Db 6111 GTAATACCATCATATATGAGAGGCTTTGGCACTGATAGTTCCCTTAATATCGGTCC 6170  
 364 ccggtatgagccttcgagatgtagaacaacgtctgtctgctgttctatcgagctacc 423

[illegible]

```

Db      7242  TACACACATGAAACATCTCATTCATGTGAGGCTCATTCATTCATCTCAACAGCAGTAATA 7301
QY      1501  ttcttcacgtgatcttt 1518
      || || | | | | |
Db      7302  TTAATAATTTTCATGAT 7319

RESULT 5
US-09-097-889-1
: Sequence 1, Application US/09097889
: Patent No. 6218117
: GENERAL INFORMATION:
: APPLICANT: Herrnstadt, Corrina
: APPLICANT: Ghosh, Soumitra S.
: APPLICANT: Davis, Robert E.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
: TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/097,889
: FILING DATE: 15-JUN-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Roseman Ph.D., Stephen J.
: REGISTRATION NUMBER: 43,058
: REFERENCE/DOCKET NUMBER: 660088.417
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6744 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-097-889-1

Query Match 23.2%, Score 388.2; DB 4; Length 6744;
Best Local Similarity 57.1%; Pred. No. 3,7e-96;
Matches 730; Conservative 19; Mismatches 517; Indels 13; Gaps 3;

QY      244  aacgacacccgctggaacgcatcagttccaccatcagtgatctcgtatcgtatcgtcttg 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2605  AACGACCCATCTACACGCTTATCTCACAGCCCATGCTTGTATATATCTCTTCATA 2684
QY      304  ggtatcccgcatcgttcgctgcgtgttctgtaactatcgtatcgctgcgtcgaactcgcgct 363
      || | || | || | || | || | || | || | || | || | || | || | || | ||
Db      2685  GTAAATACCCATCAATATGAGAGGCTTGGCAACGACTAGTTCCTCAATATATGCGTCC 2744
QY      364  cggatataggctcccgcgatagaaacaaactcgtctcgtctcgtcgttcatatgcggtacc 423
      || || || || || || || || || || || || || || || || || || || || ||
Db      2745  CCCGATATGGCGTTTCCCGCATAAACAACTAAGCGTTMTGACTCTTAACCCCTCTCTC 2804
QY      424  gcgataggcgctgcttcgctatcgcgaacggcggtgagcgatcagctggttcggggcgt 483
      || || || || || || || || || || || || || || || || || || || || ||
Db      2805  MTACTCGTGYTGGATCTGCTATAGTGGAGGCGGMCAGGAACAGCTTGAACAGT--- 2861
      || || || || || || || || || || || || || || || || || || || || ||
QY      484  ggttcggtctcgtaccgcgcgtcgtgcgaacccggaagctggtcatatgcataggacctcgc 543
      || || || || || || || || || || || || || || || || || || || || ||

```

Db	2862	-----TACCCTCCCTTTRGCAAGGAATTACTGCCAACCCMTGGAGCCTCCGTGAGACSTRACC	2915
Oy	544	atttcgcgattcgaactgtgcgtgtgactcc--tcgatactgggcgcgatactcaatgatgaac	602
Db	2916	ATTCTTCTCTTACACTACTACAGAGTGTCTCTCTTATCTTTAGGGCCATTCATTTTCATCAC	2975
Oy	603	gaactctctgaacatgacgcgccccgcgcgatagcgtgcacaagaatgcgltgttctctgtg	662
Db	2976	AMCAATTAATYAATTAATAAACCCCCCTGCACATTAACCAATATACCAAGCGCCCTTTTCGTGG	3035
Oy	663	gtcgcacttatacaagcgtctgtgcctgatccctgtctgtgcgtctgcggtcttgcgtgtgcaat	722
Db	3036	ATCCGTCCTTAATACACAGCAGTCTACTTCTCTCTATCTCTCCATCTCCATGCTGATGCGCAT	3095
Oy	723	caaccatgtctgtacccgcaccgtaactctggacagacacttctcaalcctctgtgcgcgcg	782
Db	3096	CACCTATACTACTTAACAGACCGGAMCTGCACACCCACTTCTTTTGAACCCMGCCGGAGGAGG	3155
Oy	783	tgaccgcgattctbtaccacaacatcctgtgtgtctcttbtggaccgcgcgaagtgtacatcat	842
Db	3156	AGAGCCCCATCTTAACCAACACCTATTCTGANTTTTGGGTACCCCTGAAGTTTATATCT	3215
Oy	843	caattctgcgcgcgtcttggagatacaagcgaatgtctgtctgcgaactctctcg---aaaaagc	899
Db	3216	YATTCCTACACGGCTTGCGAATAATCTCCCATTTGTACTACTACTCCGGAAAAAAGA	3275
Oy	900	gtgtcttcgtgtactctgcgatagtgtctatgtacaaatgtgtggcaatcgtgtctcggccttgt	959
Db	3276	ACCATTTTGGATACATAGTATGGCTGTGACCTATGATATATCATTTGCTTCTTAGGGTTTAT	3335
Oy	960	gcctctgtggccaccacatgttacaacglttgttaltgtctgcgtctgaaccgcgaactcctaact	1019
Db	3336	CGTGTGAGCACACCATATATTATTCAGTAGAATPAACGTAACACACAGCATATTCTTAC	3395
Oy	1020	gtctggccaacaaatgtgatgcgcgtgtgcgcgcgcgccttaagatctctctcgttgtatccgcac	1079
Db	3396	CTCCGCTACCATATATCATTCGCTATCCCAACGCGCAAGTATTTTACCTACCTGCGCAC	3455
Oy	1080	gattgtgggcgcgcctcgsattagttcaaacacgcgcgaatgtctctgtgccttbtgttattgt	1139
Db	3456	ACTCCACGGAACAAATATGAATGATCTGCGAGAGTCTGTAGCGCTTAGAGATTATATTT	3515
Oy	1140	ccgtgttcaacgltgggtgtgtgtgaacggtatcgtctgcgcgcccaagcgggtctgtgcgcgtgc	1199
Db	3516	TCTTTTACCGTAGGTAGTGCGCTGTGACTGGCATTTGATTAGCAAACTCATCATAGACATTCGT	3575
Oy	1200	atactacgaacaacctatacgttggctgggcgcacttccatacttgatgtctgcgtgtgtgcgt	1259
Db	3576	ACTACACGACACGTAACGTATTGATAGCCCACTTCCACATATGTGCTATCATATAGGAGCTGT	3635
Oy	1260	ctttgtgacacttcgcgcgtatcacttcttacaatgacgcgaagtctctgcgcgcgcgttctcc	1319
Db	3636	ATTTCGCATCATATAGAGGCTTCAATTCACATGATTTTCCCTATTCTCAGGGTACACCCATAGA	3695
Oy	1320	ggaattggagctbtgaaagcttgacactcttggaaccttcttcaactcgtgtgcgaagtcaacttct	1379
Db	3696	CCAAACTACGCCCAAAATTCATTTCGCTATCATATTATCATTCGCGGTAAATCTTAACCTTTCTT	3755
Oy	1380	ccgcgcgcgaactcttcccttggagcgtctcaaggatgacgcgcgcgttacaatogactatcccgaa	1439
Db	3756	CCCAACACACTTTTCTGGCCATATCCGGAATGCCCCAGACTTACTCGGACATTAATCCCGATGC	3815
Oy	1440	cttgcgcgtctgtgaaacaaagtctctgtctcctatgtgtgtctctcgcgtctcgtctgttct	1499
Db	3816	ATAACACCAATGAAGAAATTCCTATCATCTGTAGGCTCATTTTCATTAATTTCTTAACAGCAGTAAT	3875
Oy	1500	gtctctcaactgtgactctt 1518	
Db	3876	ATTAAATAATTTTCATTAATT 3894	





```

QY 604 accctcttgaaacatgacgagcccgatgacgctgacacaaagtccgtgtctcgtg 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 570 ACAATTATCAATATTAACACCCCTGACATACCAATTAACCAACGCCCTCTGCTCA 629
QY 664 tggattctatcatcggcgtgaccccgacgctgacgctgacgctgacgctgacg 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 TCGGCTCTAATTCACAGACTGCTACTTCTCTATCTCTCCAGTCTCTACTGCTGCAATC 689
QY 724 accatgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 ACATATACCTACTACAGACCGCAACCTCAACACACCTCTTCGACCCCGGAGAGGA 749
QY 784 gaccgacgtctgacacacacacacacacacacacacacacacacacacacacac 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 750 GACCCCAATCTATACCAACACCTATCTATCTATCTATCTATCTATCTATCTATCT 809
QY 844 attctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 900
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 810 ATCTTACAGGCTTCGGAATATCTCCATATTTGTAACCTTACTACCGGAAAGAA 869
QY 901 gctctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 960
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 870 CCATTTGATACATAGATATGCTGATGATGATGATGATGATGATGATGATGATGAT 929
QY 961 gctcgggacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 930 GTGTGACACACATATATTTACAGTAGAGATAGACGTAGACACACAGACATATTTTACC 989
QY 1021 ctgacgacacatgagtgacgacgacgacgacgacgacgacgacgacgacgacg 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 990 TCGGCTACCAATATATCATCGCTATTCACACGCGGTCAAACTATTTAGTACTGCGCACA 1049
QY 1081 atgtgggacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 CTCACGACGACATATAGAAATATCTGTCGACGTCTGACGCTGACGCTGACGATTC 1106
QY 1141 ctgtacacgctggagtgatgacgacgacgacgacgacgacgacgacgacgacgacg 1200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1107 CTTTTCACCGTAGTGGCTGACTGCGATTTGTAATAGCAACATCATACAGACATGCTA 1166
QY 1201 tatacagacacatacgtgagtgacgacgacgacgacgacgacgacgacgacgacg 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1167 CTACACGACACGACTACTGTTGATGAGCCACCTTCACATATCTCTATCAATAGAGAGCTGA 1226
QY 1261 ttctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1227 TTTGCTCATATAGAGGCTTCATCTACTGATTTCCCTATTTCTAGGCTTACACCTTAGAC 1286
QY 1321 gaacggacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1287 CAACACGACGACCAATATTCATATCATATTTGATGCGGCTGAATTAATTAATCTTCTTC 1346
QY 1381 ccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1347 CCACACGACGCTTCTGCGCTATTCGGAATGCCCGACGCTTACTGCGCTACCCGAGTCA 1406
QY 1441 ttgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1407 TACACGACATGAACATCTATCTATCTGTAGGCTCATTTCTCTAAGACAGATATA 1466
QY 1501 ttctcatcgtgacgtt 1518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1467 TTATTAATTTTCATGATT 1484

```

RESULT 8  
 US-08-998-416-564/C  
 : Sequence 564, Application US/08998416  
 : Patent No. 6239264  
 : GENERAL INFORMATION:  
 : APPLICANT: Philippesen, Peter  
 : APPLICANT: Pohlmann, Rainer

```

: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jurgen
: APPLICANT: Knechtle, Philipp
: APPLICANT: Reibschung, Corline
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESS:
: ADDRESS: No. 6239264artis Corporation
: STREET: 3054 Cornwalis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PE/5-30306/A/CCG1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 564:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 773 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG1390UP
: US-08-998-416-564

```

Query Match 8.8%; Score 147.2; DB 4; Length 773;  
 Best Local Similarity 51.5%; Pred. No. 6.9e-31;  
 Matches 369; Conservative 0; Mismatches 338; Indels 9; Gaps 1;

```

QY 250 cacctgtgaaacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 710 CAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 651
QY 310 ccgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 CCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 591
QY 370 atgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 ATATCATTTTGTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
QY 430 ggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 TTAATTTGTTTATTAACATCTATATATTAATTAATTAATTAATTAATTAATTAAT 480
QY 490 gttctgtacccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 TATCTCTCTTATCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
QY 550 ggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 TCTTTACATTTTACTACTATTTCTTCTATTAATTAATTAATTAATTAATTAATTA 360

```





TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 1138:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 719 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: PAG1692UP  
 US-08-998-416-1138

Query Match  
 Best Local Similarity 51.1%; Score 111; DB 4; Length 719;  
 Matches 261; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 1008 atccctactcgtcgtgcccacatggtgacggtgcccacgacatgaatctctc 1067  
 DB 15 ACCCTATTTTACTTACTGACTACTATATTTCTTATTTCTACTAGATTAAAGTATTAG 74  
 QY 1068 gtggatcgccacagatggtggggcggtcggttgatgtaaaatcgccagatgcttgccctt 1127  
 DB 75 TTGATTACTTAATCTTATATGTTGTTCAATTAAGATTACTTAACACCAATATATATCTATT 134  
 QY 1128 tggcttaatgctcgttcacggtggtggtgacccggtatcggtcgccacgacgag 1187  
 DB 135 ATCATTTTATTTTATTTTACTGATGATGCTTTAACTGCTGATTAAGTAAATCTATC 194  
 QY 1188 tctggacggtgacatcacacacacatgacgtggtgacacatcactatgagtc 1247  
 DB 195 ATTAGATGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 254  
 QY 1248 gctgggtgacatcgttgacatcgttgacatcgttgacatcgttgacatcgttgacatcgt 1307  
 DB 255 TTATAGTGTGATTTCTCTATTTGCTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 314  
 QY 1308 ccgacgttcctccgacgtggtgac 1367  
 DB 315 TTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 374  
 QY 1368 cgtcagcttcctccgac 1427  
 DB 375 TATTTATTTTCTTCCCTATGCTATTTCTTATGTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 434  
 QY 1428 ctatccgaagcctcgacgtggtgac 1487  
 DB 435 TTATCTGATCTATTTCTTATGCTTGAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 494  
 QY 1488 cgcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttc 1518  
 DB 495 TATATCATTTAATGATTTATTTCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 525

# RESULT 13

US-08-998-416-289  
 Sequence 289, Application US/08998416  
 Patent No. 6239264  
 GENERAL INFORMATION:  
 APPLICANT: Philippsen, Peter  
 APPLICANT: Pohlmann, Rainer  
 APPLICANT: Steiner, Sabine  
 APPLICANT: Mohr, Christine  
 APPLICANT: Wendland, Jurgen  
 APPLICANT: Knechtle, Philipp  
 APPLICANT: Reibischung, Corinne  
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII  
 TITLE OF INVENTION: AND USES THEREOF  
 NUMBER OF SEQUENCES: 1152  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwallis Road  
 CITY: Research Triangle Park  
 STATE: No. 6239264th Carolina  
 COUNTRY: USA  
 ZIP: 27709  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/998,416  
 FILING DATE: 24-DEC-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 0016/97  
 FILING DATE: 31-DEC-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: PP/5-30306/A/GC1976  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 289:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 856 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: PAG1241UP  
 US-08-998-416-289

Query Match  
 Best Local Similarity 51.1%; Score 111; DB 4; Length 856;  
 Matches 261; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 1008 atccctactcgtcgtgcccacatggtgacggtgcccacgacatgaatctctc 1067  
 DB 15 ACCCTATTTTACTTACTGACTACTATATTTCTTATTTCTACTAGATTAAAGTATTAG 74  
 QY 1068 gtggatcgccacagatggtggggcggtcggttgatgtaaaatcgccagatgcttgccctt 1127  
 DB 75 TTGATTACTTAATCTTATATGTTGTTCAATTAAGATTACTTAACACCAATATATATCTATT 134  
 QY 1128 tggcttaatgctcgttcacggtggtggtgacccggtatcggtcgccacgag 1187  
 DB 135 ATCATTTTATTTTATTTTACTGATGATGCTTTAACTGCTGATTAAGTAAATCTATC 194  
 QY 1188 tctggacggtgacatcac 1247  
 DB 195 ATTAGATGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 254  
 QY 1248 gctgggtgacatcgttgacatcgttgacatcgttgacatcgttgacatcgttgacatcgttgacatcgt 1307  
 DB 255 TTATAGTGTGATTTCTCTATTTGCTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 314  
 QY 1308 ccgacgttcctccgacgtggtgac 1367  
 DB 315 TTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 374  
 QY 1368 cgtcagcttcctccgac 1427  
 DB 375 TATTTATTTTCTTCCCTATGCTATTTCTTATGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 434  
 QY 1428 ctatccgaagcctcgacgtggtgac 1487  
 DB 435 TTATCTGATCTATTTCTTATGCTTGAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 494  
 QY 1488 cgcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttcctcgttc 1518

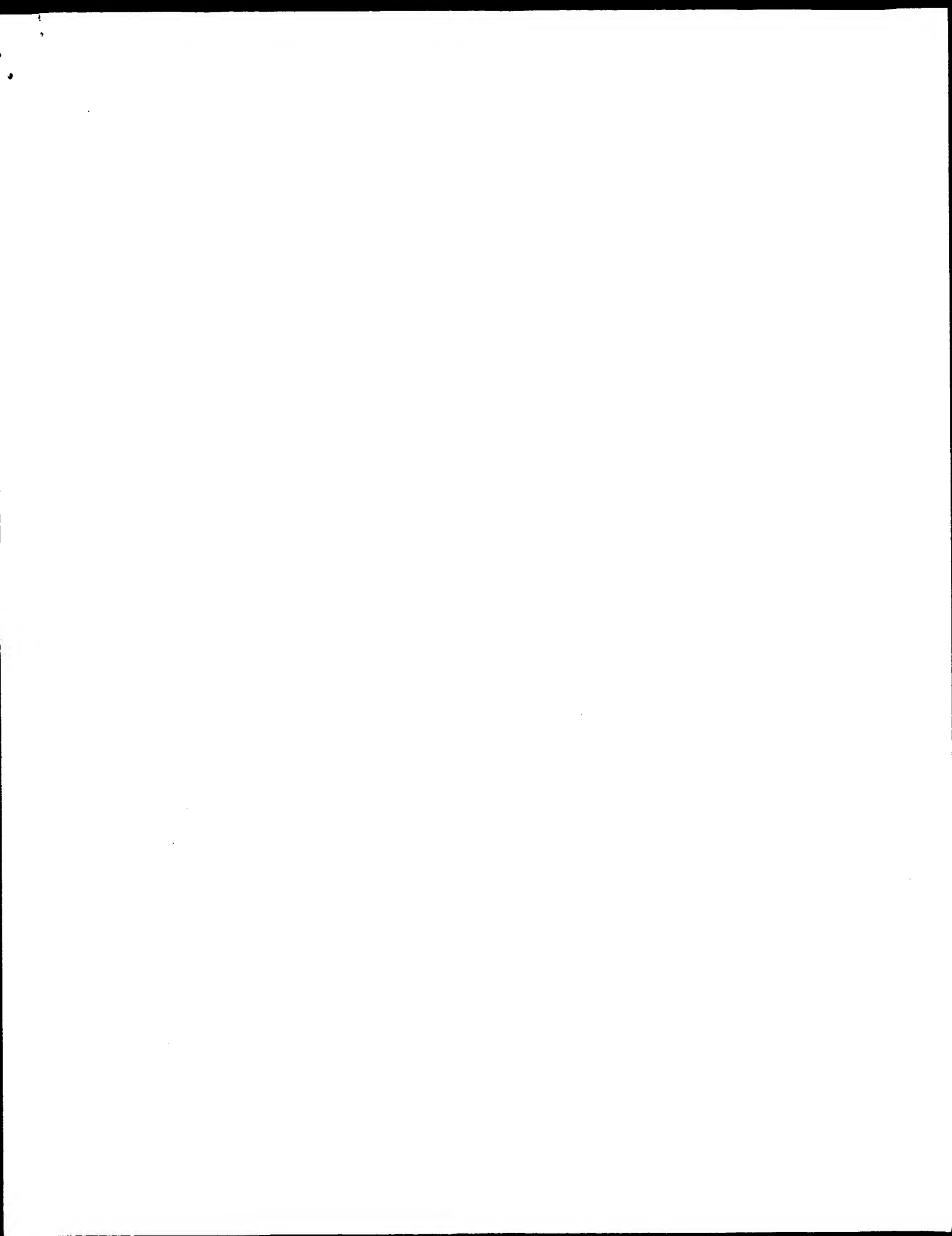


Thu Dec 27 08:21:17 2001

us-09-712-768-1.rni

---

Page 13







```

Db      86  ACGATCAATCTACAAAGCTTATTGTACCGCCAGCTTGTGAATGATTCTTTAAAG 145
Oy      305  gtatccccgaatgttcgggtgttlttgtaacatctgtagccgcgcgaatggcgctc 364
Db      146  TTATAGCTTATCAATAATTTGGAGGATTTGGCAACTGATTTGGTACATTAATAATTGGTGAC 205
Oy      365  cggatagcccttcgcgcgtatgacaacctgtcgtctcgttcgtatcgttcgcgtacgc 424
Db      206  CCGGACATAGCATTTCCACAGGATAATAACATTAAGCTTCTGACTTT-----TACC 255
Oy      425  cgaatggcggtgttcgcgtctgcacccggcggtgacggtcagctgtgttcggcggtg 484
Db      256  CCATCTTTCTTCTTCTTACTAGCTCTCAACAGTAGAGGAGGCGGACGAGACAGATGA 315
Oy      485  gtgtgtgttcgttacccgcgcgtgcgacccg--cgaaagcttgcatctgtagaaccttcgcg 543
Db      316  ACTGTGACCTCCATTAAGCTGTGTAACCTTAGCCACGCTGGGCTTGTGATTTTAGCC 375
Oy      544  atttcgggttcactgttcgggtgtcctcgtatcgtatggcgcgatcaacatgatacgc 603
Db      376  ATTTCTTCCCTCCACAGCTAGAGGTGTATCATCTTCTAGGGGCAATTAATTTCATTAAC 435
Oy      604  acctcttgaacatgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 663
Db      436  ACAATTTATTAATAAACCACGACCTATCTCAATATCAAAACCCCTCTATTGTTTGA 495
Oy      664  tgcgtcttatacagcggtgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 723
Db      496  TCTGTATAATTAACAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 555
Oy      724  accatgtctgtacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 783
Db      556  ACATAGCTTTAAACAGATGCAATCTAAACACACATCTTCGATCTCGCGGAGGTGGA 615
Oy      784  gaccgcgtctgtatacacaacatctcgtgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgt 843
Db      616  GACCAATTTATTAACCAATCTTTCTGATTTGTTGGTGTACCCGGAAGTTTACATTTTA 675
Oy      844  attcgcgcgcgtctgtgcatcatcagcatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 900
Db      676  ATTCTCCAGGCTTTGGCATATCTCCCATATTTGTAACCTTTACTCAGGTAATAAAGAA 735
Oy      901  gtctcgttaccgtgcgcgtgtctatgacatggtgtgcaatcgggtgtcgtcgtcgtcgtc 960
Db      726  CCTTTGGCTATTAAGCATGTGTGAGCTTAATTAATCTATTTGGTCTCTAGGCTTCATC 795
Oy      961  gtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
Db      796  GTTTGAGCACACACATATTTACAGTTGACTAGACGTTGATACCCGAGCTTACTTTACT 855
Oy      1021  ctgagccacatctgtgtatcgcgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
Db      856  TCTGCGCATATGATTAATGTGCAATTCACAGAGTAATAAGATTAGTTGATTAGCAACA 915
Oy      1081  atgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
Db      916  CTCGATGAGGAGAAACATCAATGATCCACGCTTACTATGAGCCCTTGGATTATTTTTC 975
Oy      1141  ctgttccacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
Db      976  CTATTTAATTCAGTGTGATTAACAGATTGTGCTAGGCCAATCTCTCACTAGACATGCTTT 1035
Oy      1201  taccagcacacttatacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
Db      1036  CTCACGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
Oy      1261  ttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
Db      1096  TTCCGATTAATTAAGGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1154
Oy      1321  gaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380
Db      1155  GACTTATGAGGAGAAATCCACTTCTCTATTTATTTGTTAGGTGCAATTAATTAACCTTCTTC 1214

```

```

Oy      1381  ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440
Db      1215  CCTCAACATTTCTCTAGGCTTATTCGGCATACGACAGCATACTGATTAACGAGACGCC 1274
Oy      1441  ttcgcgtgtgtgacaagaatctcgtcgtatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
Db      1275  TACACTACATGAAGATGCTATTTCTCTATTCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1334
Oy      1501  ttcttcacgtgtatctt 1517
Db      1335  TTAATAATGATTCATCAT 1351

```

```

RESULT 2
AUI67717 1124 bp mRNA EST 29-JAN-2001
LOCUS AUI67717 01-br-ad cDNA Oryzias latipes cDNA clone br0955, mRNA
DEFINITION
sequence.
ACCESSION AUI67717
VERSION AUI67717.1 GI:12589786
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 1124)
REFERENCE Mita, K., Ishikawa, Y. and Yamauchi, M.
Establishment of cDNA database of medaka, Oryzias latipes
Unpublished (2001)
TITLE Unpublished (2001)
JOURNAL Contact: Mita K
COMMENT Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@nirs.go.jp
method:un1-directional sequence direction:sequenced from T3 primer
(5' -> 3').

```

```

FEATURES
source Location/Qualifiers
1..1124
/organism="Oryzias latipes"
/strain="HNR"
/db_xref="taxon:8090"
/clone="br0955"
/clone_1ib="01-br-ad cDNA"
/sex="female/male mixed"
/tissue_type="brain"
/dev_stage="adult"

```

```

BASE COUNT 292 a 284 c 210 g 338 t
ORIGIN

```

```

Query Match 18.4%; Score 307.8; DB 10; Length 1124;
Best Local Similarity 59.8%; Pred. No. 5.4e-66;
Matches 535; Conservative 0; Mismatches 357; Indels 3; Gaps 1;

```

```

Oy      623  cccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 682
Db      6  CCCCCGCGACTCTCAATATACCAAAACCCCTTATTTGTTGAGCCGTAATATACAGAG 65
Oy      683  ggcgtacgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 742
Db      66  TATTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 125
Oy      743  gtaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 802
Db      126  GAACACTTAATTAACCTTTCTTGACCTTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
Oy      803  acatcgtgtgtcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 862
Db      186  ACCTTTCTGATTTCTTGGCCACCGCTGAAGTTATATTTCAATTTCTCCCGGCTTGGAA 245

```







TITLE	A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function									
JOURNAL	Genome Res. 10 (12), 2062-2069 (2000)									
MEDLINE	20568495									
COMMENT	Contact: Buerstedde JM Cellular Immunology Heinrich-pette-Institute Martinistr. 52, 20251 Hamburg, Germany Email: URL: <a href="http://genetics.hpi.uni-hamburg.de/dtd0estr.html">http://genetics.hpi.uni-hamburg.de/dtd0estr.html</a> . Location/Qualifiers									
FEATURES	source 1..821 /organism="Gallus gallus" /strain="CB" /db_xref="taxon:9031" /clone="15g5r1" /clone_lib="dkfz426" /issue_type="Bursa of Fabricius" BASE COUNT 215 a 254 c 142 g 209 t 1 others									
ORIGIN										
Query Match	17.0% Score 283.8; DB 10; Length 821;									
Best Local Similarity	62.8% ; Pred. No. 4,5e-60;									
Matches 457; Conservative	0; Mismatches 268; Indels 3; Gaps 1									
QY	516	cgaaactggtcattcatgatgacctgcgcgatttcgcggttcacctgctcggtgctccctc	575							
DB	65	CCACGCTGGCGATCAAGTAGACCTAGCCATCTTTTATTATACATTTAGACAGGTGTTTC	124							
QY	576	gatactggcgcgatacaacatgatacgaacctcttgaacatgycgcgcccgcatgac	635							
DB	125	CATTCTAGAGAGCAATCAACTTATCTACTACATCAACATCAACATMAAACCCCGCCACTGTC	184							
QY	636	gctgcacaaagycggttcttcgcgycgcgacatcctatcacgcgttcgtgtacctgct	695							
DB	185	ACAATAACCAACACCCCTATTCGTATGATCGCTTCATTAAGCCATCTACTACTCTCT	244							
QY	696	ggcgcctgcgcgtctctgcgcgtgacaaatcacacatgctcgtcgaaccgacgttaactgcgcac	755							
DB	245	CTCCTTACCCGCTCTAGCAGCTGGATTACCATATTAATTACCGACCGCAACCTTAAAC	304							
QY	756	gaacctcttcaatcctgctgctgcgcggtgacccgattctgtaccacaacatcctgtgct	815							
DB	305	CACATCTTCGACCCGCTGAGAGAGAGACCCCAATCTATACCAACACCTATTCTGATT	364							
QY	816	ctttgggcaccgcgaagtgatacatcatctctgcgcgcgttcgtgacatacgaacatg	875							
DB	365	CTTCGGTCAACCCGGAAGTTTACATCTCTCACTCTCCAGCTTGGGAATTAATTCCACGT	424							
QY	876	cgctgacaccttc--gaaaaagccgylctcgatllacatcgcgaltgltcatgcaat	932							
DB	425	AGTAGACATCTATGTCAGGAAMAAAAMAAACCATTCGGATACATAGCAATATGTCAGCCAT	484							
QY	933	ggtggcaatcgtgtctcgtgacctgtgctgttcgttcggcgacacacatgatacgcgttgat	992							
DB	485	ACTGCTAATCGAATTCCTTGGCTTCATTGTATAGACCCACCATATATATACAGTCGGAAT	544							
QY	993	gtcgcgcgaaccagaacatcactcaatgctcgcgcgaacatgltgacgcgtgcgcgacg	1052							
DB	545	GGACGTAGACACCCGGAGCCTACTTTATATCAGGCACACATATATCATGCCATCCCAATGG	604							
QY	1053	catlaagatcttcgtgtagacgcacagatgtaggcgcgtcggttgatgtaaatgc	1112							
DB	605	TATTAAAGTCTTCAGCTAGATAGCAACCCGTCACGGAGGAACAATTAAAGACACCCCCC	664							
QY	1113	gatgcctcgtgaccttgaccttattgttcctgctcacccgtgggtggtgttgaccgtatcgt	1172							
DB	665	TATGCTATGTGCCCCATAGATTATCTTCCTTCCTTACATATGTGGAGGCGCTTAACGGGAATCGT	724							
QY	1173	gtgcgcgaacgaggtcttgacgcgtgataatacgaacacatatacgtgtgtagcgaact	1232							
DB	725	CTTTCCTACTCATCACTAGATATATGTCCTTCATAGACACCTACTAGATATGAGCGCCACTT	784							

	QY	1233	coatatg	1240
	Db	785	CCACTATG	792

RESULT	8			
LOCUS	AJ393803	810 bp	mRNA	EST
DEFINITION	AJ393803 dktf426 Gallus gallus cDNA clone 16o2r1,			25-JAN-2001
ACCESSION	AJ393803			
VERSION	AJ393803.1	GI:7123057		
KEYWORDS	EST.			
SOURCE	chicken.			
ORGANISM	Gallus gallus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
	Phasianinae; Gallus.			
REFERENCE	1 (bases 1 to 810)			
AUTHORS	Abdrakmanov,I., Lodgyn,D., Gerolt,P., Arakawa,H., Law,A., Plachy			
TITLE	J., Korn,B. and Buerstedde,J.M. A large database of chicken burial ESTs as a resource for the			
	analysis of vertebrate gene function			
JOURNAL	Genome Res. 10 (12), 2062-2069 (2000)			
MEDLINE	20568495			
COMMENT	Contact: Buerstedde JM Cellular Immunology Heinrich-Pette-Institute Martinistr. 52, 20251 Hamburg, Germany Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.			
FEATURES	Location/Qualifiers			
source	1..810			
	/organism="Gallus gallus"			
	/strain="Cb"			
	/db_xref="taxon:9031"			
	/clone="16o2r1"			
	/clone_lib="dktf426"			
	/issue_type="Bursa of Fabricius"			
BASE COUNT	216 a 254 c 130 g 209 t			1 others
ORIGIN				

Query Match	16.8%;	Score 281.8;	DB 10;	Length 810;
Best Local Similarity	60.6%;	Pred. No. 1.4e-59;		
Matches 479;	Conservative	0;	Mismatches 308;	Indels 3; Gaps 1

QY	613	aacatgcgcgcccccgagcgtacgcgtgcgaagaagtgcgcttgcttcgctgatcatctt	672
Db	1	ANCATMAAACCCCCCGCAGTGCACATPACGAACACCCTATTCTGTATGATTCGGTCTC	60
QY	673	atcaaggcttgtactgacctgtcgtcgtcgctgcggcttcgtgcgtgtgacaaccacatcg	732
Db	61	ATTACTGCCATCTCTACTACTCTCTCTCTTAACCGTGCTCTGACAGCTGGATTACCATCTA	120
QY	733	ctagcgcgcgcgtacttcgcgaacgaccttcctaaccgtcgtgcggcggttgaccgat	792
Db	121	CTTTACGACCGCACACTTAAACCAACAATTCTTGACCCGACGtGAGAGAGACCAATC	180
QY	793	ctgtaccacaacatccctgttgtctctttggtgacccggaaagtgtacalcacattctgccc	852
Db	181	CTATTACCAACACCTATTCTGATTTCTTCGGHCACCCGGAAGTTTACATCTCTCCCA	240
QY	853	gaccttggcatcacatcagccatgcgtgcgttcgaacttctc---gaaaagccggtctcgt	909
Db	241	GGTTTTGGAAATAATTTTCCACGTAGTAGCACTATCTATGACAGAAGAAAAAGAACCATTC	300
QY	910	tacctgcgcgtgtgtcatgtaaatcgtgtggccaatcggtgttccttggtcttgcgtctg	969
Db	301	TACATAGGAATAGTCTGAGCCATACCTCTCATATGGGATCTCTTGCTTCATTTGATGAGCC	360
QY	970	caccacatgtacacgcttgttatgtcgcgtgtgaccagaaatcactaacatgcgtgcacc	1029
Db	361	CACCATTATTTTACAGTGGGAATGGAGCGTGGACACCCGAGACCTTATCATCAGCCACA	420

Query Match	16.7%	Score 279.6	DB 10	Length 866
Best Local Similarity	60.0%	Pied 5.1e-59		
Matches 502	Conservative	0	Mismatches 330	Indels 5
			Gaps	
Qy	623	ccccgcgcataagcgcgcgcacaaagtcgcgtgtgttcctcgtgctgcgatctttatacagcctt	682	
Db	9	ccccgcgcactgtacacatattccaaacaccccttatttcgatgatgctgcgcgttcattatgcgca	68	

RESULT	10
AJ393270	
LOCUS	AJ393270 804 bp mRNA EST 25-JAN-2001
DEFINITION	AJ393270 dkffz426 Gallus gallus cDNA clone 14o7r1, mRNA sequence.
ACCESSION	AJ393270
VERSION	AJ393270.1 GI:7121964
KEYWORDS	EST.
SOURCE	chicken.
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 804)
AUTHORS	Abdulkhanov,I., Lodygin,D., Gerotch,P., Arakawa,H., Law,A., Plachy .J., Korn,B. and Buerstedde,J.M.
TITLE	A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function
JOURNAL	Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE	20568495
COMMENT	Contact: Buerstedde JM



Cellular Immunology  
 Heinrich-Pette-Institute  
 Martinstr. 52, 20251 Hamburg, Germany  
 Email: URL: http://genetics.hpi.uni-hamburg.de/dl40est.html.  
 Location/Qualifiers

FEATURES  
 source 1..804  
 /organism="Gallus gallus"  
 /strain="CB"  
 /db\_xref="taxon:9031"  
 /clone="14071"  
 /clone.lib="dkfz426"  
 /tissue.type="Bursa of Fabricius"

BASE COUNT 213 a 252 c 129 g 210 t  
 ORIGIN

Query Match 16.7%; Score 279.2; DB 10; Length 804;  
 Best Local Similarity 60.8%; Pred. No. 6.2e-59;  
 Matches 474; Conservative 0; Mismatches 303; Indels 3; Gaps 1;

623 ccccgagatgacgctgcacaaagtcggtgttcctcggtggtcgtatctatcaagctt 682  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 8 ccccgagatgacgctgcacaaagtcggtgttcctcggtggtcgtatctatcaagctt 67  
 683 ggtgtagctgctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 742  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 68 tccctactact 127  
 743 gtaactcgacgacgact 802  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 128 gcaacgttaccac 187  
 803 acatctgtgttct 862  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 188 acctatctatct 247  
 863 tcatcagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 919  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 248 taaatttccacgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 307  
 920 tggctatgcaatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 979  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 308 tagctgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 367  
 980 acacgctgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1039  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 368 tccagctgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 427  
 1040 cgtgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1099  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 428 ccaatccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 487  
 1100 agtcaaatcgcgagctgctggtggtggtggtggtggtggtggtggtggtggtggtg 1159  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 488 aatagagccccct 547  
 1160 tggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1219  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 548 taaaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 607  
 1220 tggctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1279  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 608 tagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 667  
 1280 tctactttacatgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1339  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 668 ttaccacactgatttccct 727  
 1340 actctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1399  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 728 atttggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 787

## RESULT 11

AUI69837

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE  
 Mita,K., Ishikawa,Y. and Yamuchi,M.  
 Establishment of cDNA database of medaka, *Oryzias latipes*  
 Unpublished (2001)  
 Contact: Mita K  
 Genome Research Group  
 National Institute of Radiological Sciences  
 Anagawa 4-9-1, Inage, Chiba 263-8555, Japan  
 Email: kmita@nirs.go.jp  
 method:unl-directional sequence direction:sequenced from T3 primer  
 (5' -> 3').

## FEATURES

source  
 location/Qualifiers  
 1..809  
 /organism="Oryzias latipes"  
 /strain="HNI"  
 /db\_xref="taxon:8090"  
 /clone="br5288"  
 /clone.lib="OI-br-ad cdna"  
 /sex="female/male mixed"  
 /tissue.type="brain"  
 /dev\_stage="adult"

BASE COUNT 203 a 193 c 153 g 259 t 1 others  
 ORIGIN

Query Match 16.5%; Score 276.6; DB 10; Length 809;  
 Best Local Similarity 60.8%; Pred. No. 2.8e-58;

Matches 486; Conservative 0; Mismatches 309; Indels 4; Gaps 2;

661 tggctatcttatacagcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgct 720  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 8 tggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 67  
 721 atccacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 780  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 68 atcactatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 127  
 781 ggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 128 gggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 187  
 841 atcactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 897  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 188 ctatattctcccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 247  
 898 ccggtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 957  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 248 gaaacttggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 307  
 958 gtcgctggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1017  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 308 attgtttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 367  
 1018 atgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1077  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 368 acattgcaacacatatttctgctgctgctgctgctgctgctgctgctgctgctgctg 427  
 1078 acgatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1137  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||







```

Db 254 GCTATATAGCAATAGATAGAGCAATATATGCTTATGAGCTTATGATGAG 313
Qy 968 cgcacacacatgacacgctgtgatacgtgacacgaacatctactcaacgctgca 1027
Db 314 CCACACCAATATTCACAGTATGATGATGATGATGATGATGATGATGATGATG 373
Qy 1028 ccatggtgacgctgacgacgacgacgacgacgacgacgacgacgacgacgacg 1087
Db 374 CTATATATATGCAATATTCACGAGTGTCAATATTTAGCTTACTTGAACCTTACAG 433
Qy 1088 ggcgcgcgcgtgtgacgacgacgacgacgacgacgacgacgacgacgacgacg 1147
Db 434 GAGGTATATATATATATATATATATATATATATATATATATATATATATATAT 493
Qy 1148 cgcgtggtggtgtgacgacgacgacgacgacgacgacgacgacgacgacgacg 1207
Db 494 CATTGTGTGTCTTACCGGATTTGTTTATCCACATCATCTCCCTGACATGCTTACAG 553
Qy 1208 aacacattacgctgtgacgacgacgacgacgacgacgacgacgacgacgacgacg 1267
Db 554 ATACATATATATATATATATATATATATATATATATATATATATATATATATAT 613
Qy 1268 tcttcgcgcgtgtgacgacgacgacgacgacgacgacgacgacgacgacgacg 1327
Db 614 TCATATACAGATTTGTTCACTGATTTCCATTTATTTTACAGGCTTACATGATGACAT 673
Qy 1328 ctgcacacgctgtgacgacgacgacgacgacgacgacgacgacgacgacgacg 1387
Db 674 GAGCAAAAGCCCATCTGCGCATCATATTTGTAGAGATTAACATATATCTTCCCTCAC 733
Qy 1388 actctctgtgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1446
Db 734 ATTTCTGTGGCTTTTTCAGAAATACACGACGCTACTACTACAGATGCTTACAC 793
Qy 1447 cgtggaacacgctgtgacgacgacgacgacgacgacgacgacgacgacgacg 1466
Db 794 AATGAACACGCTGTCTTTC 813

RESULT 15
LOCUS BG783541 822 bp mRNA EST 20-MAY-2001
DEFINITION SEAMC003498 Sea urchin primary mesenchyme cell cDNA library
Strongylocentrotus purpuratus cDNA clone PC_0018_B1_E11_MR_5', mRNA
sequence.
ACCESSION BG783541
VERSION BG783541.1 GI:14154554
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinocystia;
Echinoidea; Echinoidea; Echinocystia; Echinocystia; Echinocystia;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 822)
AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
TITLE A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
JOURNAL Development 128 (2001) In press
COMMENT Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.

FEATURES
source
1..822
location/Qualifiers
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="PC_0018_B1_E11_MR"
/library="Sea urchin primary mesenchyme cell cDNA

```

```

/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/!db_host="E.coli"
/!note="Vector: pSPOR1; Site.1: NotI; Site.2: SalI; oligo
dt priming from poly A+ RNA, directionally cloned"
BASE COUNT 169 a 203 c 166 g 263 t 1 others
ORIGIN

Query Match 16.0%; Score 268.6; DB 11; Length 822;
Best Local Similarity 61.6%; Pred. No. 2.7e-56;
Matches 446; Conservative 0; Mismatches 275; Indels 3; Gaps 1;

Qy 518 aagcgcgcatcgtgacgacgacgacgacgacgacgacgacgacgacgacgacg 577
Db 89 ACGCCGAGGCGCTGTATATATATATATATATATATATATATATATATATATATAT 148
Qy 578 tcatgagcgcacacacacacacacacacacacacacacacacacacacacacac 637
Db 149 TCTTGCCCTCAATTAATTTATACACATTTATATATATATATATATATATATATAT 208
Qy 638 tgcacaaagtgccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 697
Db 209 TTGATCGTCCTTCTTATTCCTGATTCGCTTGTGACATGCTTGTGCTCTCTCTT 268
Qy 698 cgcgcgcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 757
Db 269 CTCTTCCAGTATTTAGCAGACGACATTTACAAATGCTTCTCAGATGCAATTAACAC 328
Qy 758 ccttctcaatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 817
Db 329 CTTCCTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 388
Qy 818 ttggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 877
Db 389 TTGACACCCCGAGGTGTATTTCTTATCTTACCGGATTTGTATGATCTCACACGTT 448
Qy 878 tgcgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 934
Db 449 TAGCTCACTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 508
Qy 935 tggcaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 994
Db 509 TTGCAATATGAGATTTTATGATTTCTTGTGCGCCACATATGTTTACAGTGGATGG 568
Qy 995 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1054
Db 569 ATGTTGATPACAGACGACATCTTCACTGCGCCACAGATGATTTATGCTTCCAAACAG 628
Qy 1055 ttaagatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1114
Db 629 TTATNGTTTTCAGATGATGACACGCTTCCAGGCTTCAATCTTCAATGAGAAACCTCT 688
Qy 1115 tgcctcggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1174
Db 689 TATATATGAGCTTGTGATTTGTATTTATTTATTTATTTATTTATTTATTTATTT 748
Qy 1175 tggcccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1234
Db 749 TTGCCAATTTCTTCATTTAGCGTTGTCTTCAATTTACTTCAATTTAGCGTTAGCTTTC 808
Qy 1235 atta 1238
Db 809 ACTA 812

Search completed: December 22, 2001, 08:48:25
Job time: 2446 sec

```

Thu Dec 27 08:21:24 2001

us-09-712-768-1.rst

---

Page 12